4.6. Synchrotron small-angle x-ray scattering (SAXS)—perspectives of ML

Peter Fratzl
Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

Status

X-ray scattering and diffraction pertain to a major set of techniques to characterize the structure of materials at the nanoscale. SAXS, in particular, has been developed in the 1950s to resolve structures in the size range 1–100 nm [194]. Despite the development of electron microscopes some years later, it remained an important technique, mostly because x-rays are less strongly absorbed than electrons, which allows for in-operando experiments, studying the effect of physical stimuli, such as temperature, pH or humidity on material structure. A strong boost in the use of small-angle scattering came with the availability of synchrotron radiation that improved the time resolution of in-operando experiments, but also opened to possibility to transform SAXS into a multiscale imaging tool. In this approach, the general idea is that nanoscale information is extracted from analysing the scattering patterns, while mapping of the specimens provides the information at the microscale (see figure 23). The first attempts with SAXS-based imaging go back to the 1990s [195]. This evolved until the development of SAXS tomography which yields six-dimensional data: three dimensions in real space through scanning and rotating the specimen (typically with micrometre resolution), as well as three additional dimensions from the scattering patterns within each voxel (containing nanoscale information) [196, 197].

The enormous advance in the brilliance of x-ray beams, as well as in x-ray optics enables not only the collection of multidimensional SAXS-tomography data but also the measurement of massive numbers of specimens even within short times.

Current and future challenges

These advances upstream of the specimen in the experiment, however, lead to new challenges downstream of the specimen, linked to the treatment and the evaluation of massive amounts of data. A schematic of the workflow in a SAXS measurement is shown in figure 24. The traditional way of conducting such an experiment would be the path symbolized by (A) and (B) in this figure. (A) represents specimen preparation and the experiment planning and (B) the data collection. These data would then be brought back from the synchrotron experiment for treatment and analysis. However, with the increased speed of data collection, a general challenge in this approach resides in the fact that the experimentalist is essentially blind without some capabilities of data diagnostics. This requires elementary pre-analysis of the data to see whether a modification of the beamline setup could improve the experiment. Recognizing this, software packages involving fast data diagnostics were developed, an example being DPDAK, an open code software introduced at the BESSY and the DESY synchrotrons (in Berlin and Hamburg, respectively) [199].

With the amount of data collected in each beamtime session increasing continuously over the years, a number of additional challenges appear from the fact that manual data treatment becomes impossible. This applies to the cleaning of data (such as denoising, background subtraction, image reconstruction, normalization, etc) and even more to the data analysis, which in SAXS often involves data fitting. These steps are indicated by the arrows (D) and (E) in figure 24.

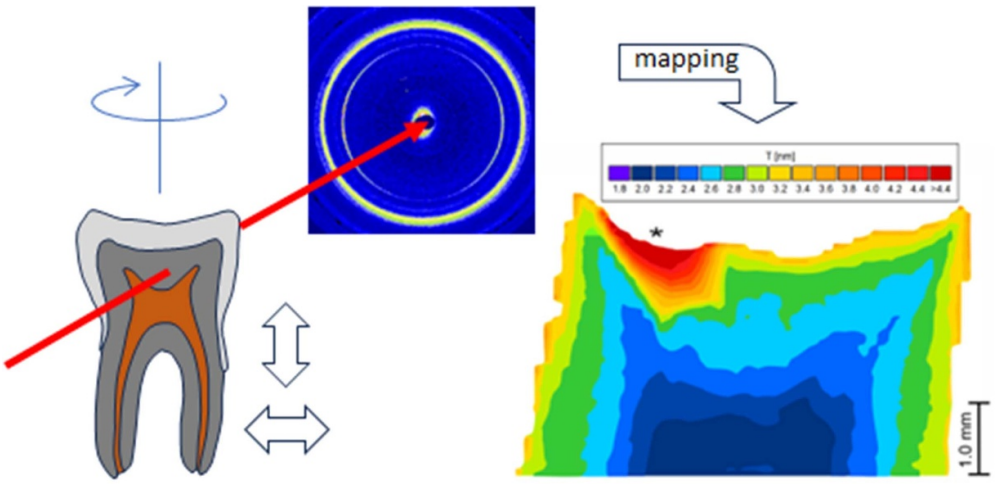

Figure 23. Principle of scanning-SAXS imaging. The specimen (for example a tooth section) is scanned across the x-ray beam with a diameter between tens of nanometres and several micrometres. Parameters extracted from the scattering patterns can then be mapped with a resolution corresponding to the x-ray beam diameter. In the figure, this is the thickness of mineral particles in dentin (the star indicates an area with a caries lesion). Reprinted from [198], Copyright (2010), with permission from Elsevier.

Advances in science and technology to meet challenges

Especially in SAXS tomography experiments, radiation damage should not be underestimated, since every specimen position will be hit several times by an intensive x-ray beam due to the required rotation of the specimen around multiple axes [197]. A typical strategy is then to reduce the irradiation time, which inevitably increases the noise in the data. To avoid problems with this noise in the 6D data reconstruction after the measurements, Zhou and coworkers propose a ML algorithm for the denoising of scattering data [200]. This approach facilitates step (D) in the diagram of figure 24.

The reconstruction of SAXS tomography data is equally challenging due to their high dimensionality. A possible traditional approach consists in calculating invariants of the SAXS data before reconstruction, which replaces the three-dimensional SAXS data by scalars that can be reconstructed much more efficiently [201]. SAXS invariants are useful, since they contain information about volume and surface of nano-size objects in the specimen [194] and allow, for example, the calculation of particle sizes in bone or dentin [195, 198, 201]. In the last few years, ML approaches are being developed for tomographic data reconstruction. Omori and coworkers review these developments for tomography using SAXS but also x-ray diffraction and other modalities [202]. While these advances relate to step (D) in figure 24, the review also addresses ML approaches for segmentation and analysis of the reconstructed data [202] (step (E) in figure 24).

Once data are reconstructed, every voxel in SAXS tomography data contains a scattering pattern to be analysed. This means a massive effort for data analysis (step (E) in figure 24) after reconstruction. Similar numbers of SAXS patterns need to be analysed in other situations, for example when material structures are studied as function of physical parameters (temperature, pressure, pH, humidity, etc) in multiple measurements. A recent review by Anker and coworkers addresses ML approaches to analyse a range of synchrotron-based experiment data,

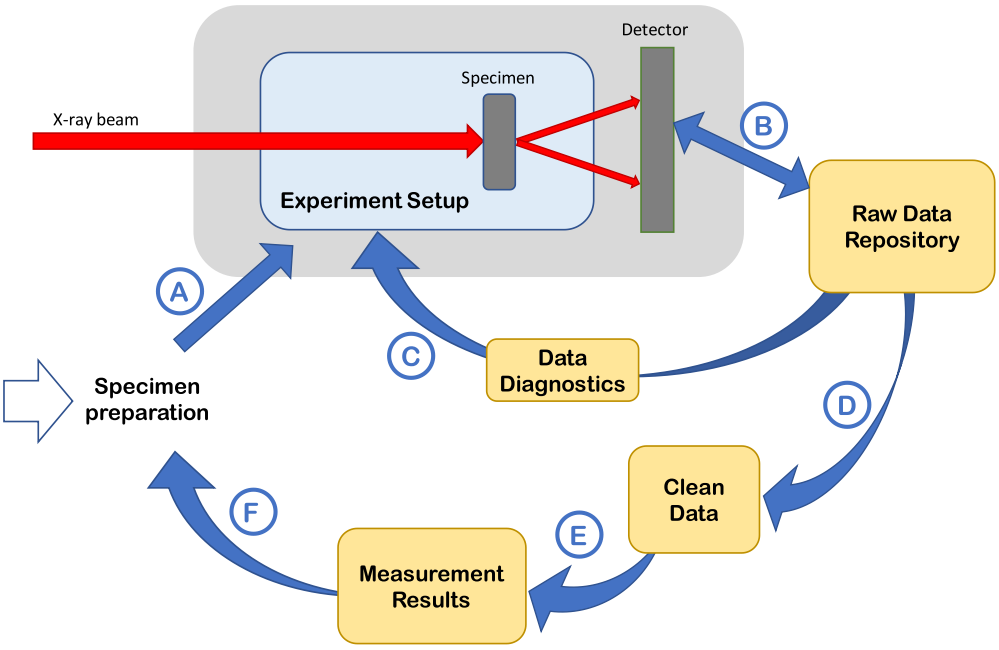

Figure 24. Schematic workflow of a small-angle x-ray scattering experiment. The traditional approach would be characterized by the arrows (A) (experiment planning to define the experiment setup) and (B) (data collection). With increasing data rates, several feedback loops involving machine learning are beginning to improve the quality and speed of the experiment: (C) is a readjustment of the experiment setup based on rapid data diagnostics, (D) is data reduction and denoising, (E) is data analysis and (F) automatic material synthesis based on the measurement results.

including SAXS but also powder diffraction, pair distribution function, inelastic neutron scattering and x-ray absorption spectroscopy data. While the traditional approach would be to fit a physical model to the data, supervised ML can be used to train a model for the prediction of structure based on data, but also to predict the scattering data based on a known structure and also to predict parameters based on some physical understanding of the system [203]. In another recent work [204], a ML-based analysis of SAXS data is proposed, which is based on Gaussian RFs that avoids the common model fitting of the data.

The approaches discussed until now are improving workflows in nearly all steps of SAXS experimentation (step (C)–(E) in figure 24). A last step (F) potentially closes the loop towards a fully automatized experimentation. This challenge is currently being taken up under the label of Autonomous Experimentation. Beaucage and Martin report on the development of an open liquid handling platform for autonomous formulation and x-ray scattering [205]. Yager and coworkers review this new paradigm and show how autonomous x-ray scattering can enhance efficiency and help discover new materials [206].

Concluding remarks

SAXS is an old method that is currently seeing an enormous increase in activity due to highly brilliant x-ray sources, more performant x-ray optics and—most recently—rapid progress in the treatment and the analysis of large amounts of data. As discussed above, several approaches

have been developed addressing some of the steps in the workflow sketched in figure 23 through ML, but there are many more opportunities for applying such methods. Faster and, therefore, more effective tools for online data diagnostics based on ML during the experiment could bring a major improvement. Indeed, this has the potential to significantly reduce measurement times and radiation damage on sensitive specimens by allowing dynamic experiment planning. Moreover, better automatic tools for data cleaning, noise reduction, as well as to correct for background and instrumental resolution will be essential for high-throughput experiments or tomographic measurements. Finally, there are further needs for combining physical models with ML methods in data fitting, a development which may require a wide-spread effort in training relevant models for a variety of material classes. In conclusion, ML approaches have an important role to play in many areas of synchrotron x-ray scattering and the developments in this direction have only just begun.

Acknowledgment

The author thanks BiGmax, the Max Planck Society's Research Network on Big-Data-Driven Materials-Science, for support and stimulating interdisciplinary interactions with members of the consortium.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Peter Benner https://orcid.org/0000-0003-3362-4103
Tristan Bereau https://orcid.org/0000-0001-9945-1271
Volker Blum https://orcid.org/0000-0001-8660-7230
Mario Boley https://orcid.org/0000-0002-0704-4968
Christian Carbogno https://orcid.org/0000-0003-0635-8364
C Richard A Catlow https://orcid.org/0000-0002-1341-1541
Sebastian Eibl https://orcid.org/0000-0002-1069-2720
Lucas Foppa https://orcid.org/0000-0003-3002-062X
Christoph Freysoldt https://orcid.org/0000-0002-7896-3478
Baptiste Gault https://orcid.org/0000-0002-4934-0458
Pawan Goyal https://orcid.org/0000-0003-3072-7780
Lara Kabalan https://orcid.org/0000-0001-5715-4332
Petr Karpov https://orcid.org/0000-0003-1388-9841
Christoph T. Koch https://orcid.org/0000-0002-3984-1523
Sebastian Kokott https://orcid.org/0000-0003-1066-6909
Igor Kowalec https://orcid.org/0000-0002-9470-1275
Kurt Kremer https://orcid.org/0000-0003-1842-9369
Andreas Leitherer https://orcid.org/0000-0001-7747-4122
Yue Li https://orcid.org/0000-0003-3377-6676
Christian H Liebscher https://orcid.org/0000-0001-8620-4597
Andrew J Logsdail https://orcid.org/0000-0002-2277-415X

Felix Luong https://orcid.org/0000-0001-7821-295X
Andreas Marek https://orcid.org/0000-0001-5403-7528
Jaber R Mianroodi https://orcid.org/0000-0003-4778-3260
Jörg Neugebauer https://orcid.org/0000-0002-7903-2472
Zongrui Pei https://orcid.org/0000-0003-0748-4629
Thomas A R Purcell https://orcid.org/0000-0003-4564-7206
Dierk Raabe https://orcid.org/0000-0003-0194-6124
Markus Rampp https://orcid.org/0000-0001-8177-8698
Mariana Rossi https://orcid.org/0000-0002-3552-0677
Jan-Michael Rost https://orcid.org/0000-0002-8306-1743
Ulf Saalmann https://orcid.org/0000-0003-3208-8273
Marcel Schloz https://orcid.org/0000-0001-6295-1715
Annette Trunschke https://orcid.org/0000-0003-2869-0181
Ye Wei https://orcid.org/0000-0003-1965-2298
R Patrick Xian https://orcid.org/0000-0001-9895-6956
Matthias Scheffler https://orcid.org/0000-0002-1280-9873

References

[1] The concept of a fourth paradigm was probably first discussed by J Gray at a workshop on January 11, 2007 before he went missing at the Pacific on January 28, 2007 Hey T, Tansley S and Tolle K (eds) 2009 *The Fourth Paradigm, Data Intensive Discovery* (Microsoft Research)

[2] Slater J C 1937 Wave functions in a periodic potential *Phys. Rev.* **51** 846
Slater J C 1953 An augmented plane wave method for the periodic potential problem *Phys. Rev.* **92** 603
Slater J C 1965 *Quantum Theory of Molecules and Solids, Symmetry and Energy Bands in Crystals* vol 2 (McGraw-Hill)
Slater J C 1967 *Quantum Theory of Molecules and Solids, Insulators, Semiconductors and Metals* vol 3 (McGraw-Hill)
Slater J C and Johnson K H 1972 Self-consistent-field Xα cluster method for polyatomic molecules and solids *Phys. Rev.* B **5** 844

[3] Hohenberg P and Kohn W 1964 Inhomogeneous electron gas *Phys. Rev.* **136** B864

[4] Metropolis N, Rosenbluth A W, Rosenbluth M N and Teller E 1953 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087

[5] Alder B J and Wainwright T E 1958 Molecular dynamics by electronic computers *Int. Symp. on Transport Processes in Statistical Mechanics* ed I Prigogine (Wiley) pp 97–131
Alder B J and Wainwright T E 1962 Phase transition in elastic disks *Phys. Rev.* **127** 359–361
Alder B J and Wainwright T E 1970 Decay of velocity autocorrelation function *Phys. Rev.* A **1** 18–2

[6] Rahman A 1964 Correlations in the motion of atoms in liquid argon *Phys. Rev.* **136** A405–11

[7] Agrawal A and Choudhary A 2016 Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science *APL Mater.* **4** 053208

[8] Draxl C and Scheffler M 2020 Big data-driven materials science and its FAIR data infrastructure *Handbook of Materials Modeling* ed W Andreoni and S Yip (Springer)

[9] Foppa L *et al* 2021 Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence *MRS Bull.* **46** 1016–26

[10] Raabe D, Mianroodi J R and Neugebauer J 2023 Accelerating the design of compositionally complex materials via physics-informed artificial intelligence *Nat. Comput. Sci.* **3** 198–209

[11] Rao Z *et al* 2022 Machine learning–enabled high-entropy alloy discovery *Science* **378** 78–85

[12] Sutton C, Boley M, Ghiringhelli L M, Rupp M, Vreeken J and Scheffler M 2020 Identifying domains of applicability of machine learning models for materials science *Nat. Commun.* **11** 4428

[13] Scheffler M *et al* 2022 FAIR data enabling new horizons for materials research *Nature* **604** 635–42

[14] Schmidt J, Marques M R, Botti S and Marques M A 2019 Recent advances and applications of machine learning in solid-state materials science *npj Comput. Mater.* **5** 83
[15] Donoho D 2017 50 years of data science *J. Comput. Graph. Stat.* **26** 745–66
[16] Sutton C, Ghiringhelli L M, Yamamoto T, Lysogorskiy Y, Blumenthal L, Hammerschmidt T, Golebiowski J R, Liu X, Ziletti A and Scheffler M 2019 Crowd-sourcing materials-science challenges with the NOMAD 2018 Kaggle competition *npj Comput. Mater.* **5** 111
[17] Lookman T, Balachandran P V, Xue D and Yuan R 2019 Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design *npj Comput. Mater.* **5** 21
[18] Shahriari B, Swersky K, Wang Z, Adams R P and de Freitas N 2016 Taking the human out of the loop: a review of Bayesian optimization *Proc. IEEE* **104** 148–75
[19] Zhan D and Xing H 2020 Expected improvement for expensive optimization: a review *J. Glob. Optim.* **78** 507–44
[20] De Ath G, Everson R M, Rahat A A and Fieldsend J E 2021 Greed is good: exploration and exploitation trade-offs in Bayesian optimisation *ACM Trans. Evol. Learn. Optim.* **1** 1–27
[21] Biau G and Scornet E 2016 A random forest guided tour *Test* **25** 197–227
[22] Efron B 1979 Bootstrap methods: another look at the jackknife *Ann. Stat.* **7** 1–26
[23] Behler J 2021 Four generations of high-dimensional neural network potentials *Chem. Rev.* **121** 10037–72
[24] Deringer V L, Bartók A P, Bernstein N, Wilkins D M, Ceriotti M and Csányi G 2021 Gaussian process regression for materials and molecules *Chem. Rev.* **121** 10073–141
[25] de Pablo J J *et al* 2019 New frontiers for the materials genome initiative *npj Comput. Mater.* **5** 41
[26] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403
[27] Musil F, Willatt M J, Langovoy M A and Ceriotti M 2019 Fast and accurate uncertainty estimation in chemical machine learning *J. Chem. Theory. Comput.* **15** 906–15
[28] Jeong W, Yoo D, Lee K, Jung J and Han S 2020 Efficient atomic-resolution uncertainty estimation for neural network potentials using a replica ensemble *J. Chem. Phys. Lett.* **11** 6090–6
[29] Hirschfeld L, Swanson K, Yang K, Barzilay R and Coley C 2020 Uncertainty quantification using neural networks for molecular property prediction *J. Chem. Inf. Modeling* **60** 3770–80
[30] Kahle L and Zipoli F 2022 Quality of uncertainty estimates from neural network potential ensembles *Phys. Rev.* E **105** 015311
[31] Tan A R, Urata S, Goldman S, Dietschreit J C and Gómez-Bombarelli R 2023 Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles (arXiv:2305.01754)
[32] Scalia G, Grambow C A, Pernici B, Li Y and Green W H 2020 Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction *J. Chem. Inf. Modeling* **60** 2697–717
[33] Jinnouchi R, Karsai F and Kresse G 2019 On-the-fly machine learning force field generation: application to melting points *Phys. Rev.* B **100** 014105
[34] Palmer G, Du S, Politowicz A, Emory J P, Yang X, Gautam A, Gupta G, Li Z, Jacobs R and Morgan D 2022 Calibration after bootstrap for accurate uncertainty quantification in regression models *npj Comput. Mater.* **8** 115
[35] Raimbault N, Grisafi A, Ceriotti M and Rossi M 2019 Using Gaussian process regression to simulate the vibrational Raman spectra of molecular crystals *New J. Phys.* **21** 105001
[36] Wrobel S 1997 An algorithm for multi-relational discovery of subgroups *European Conf. on Principles of Data Mining and Knowledge Discovery* (*Trondheim, Norway*) pp 78–87
[37] Friedman J H and Fisher N I 1999 Bump hunting in high-dimensional data *Stat. Comput.* **9** 123–44
[38] Goldsmith B R, Boley M, Vreeken J, Scheffler M and Ghiringhelli L M 2017 Uncovering structure-property relationships of materials by subgroup discovery *New J. Phys.* **19** 013031
[39] Foppa L, Sutton C, Ghiringhelli L M, De S, Löser P, Schunk S A, Schäfer A and Scheffler M 2022 Learning design rules for selective oxidation catalysts from high-throughput experimentation and artificial intelligence *ACS Catal.* **12** 2223–32
[40] Foppa L and Ghiringhelli L M 2022 Identifying outstanding transition-metal-alloy heterogeneous catalysts for the oxygen reduction and evolution reactions via subgroup discovery *Top. Catal.* **65** 196–206
[41] Grosskreutz H, Rüping S and Wrobel S 2008 Tight optimistic estimates for fast subgroup discovery *Machine Learning and Knowledge Discovery in Databases* (*Antwerp, Belgium*) pp 440–56

[42] Nguyen H V and Vreeken J 2015 Non-parametric Jensen-Shannon divergence *Machine Learning and Knowledge Discovery in Databases: European Conf.* (*Porto, Portugal*) pp 173–89
[43] Mazheika A, Wang Y-G, Valero R, Viñes F, Illas F, Ghiringhelli L M, Levchenko S V and Scheffler M 2022 Artificial-intelligence-driven discovery of catalyst genes with application to CO_2 activation on semiconductor oxides *Nat. Commun.* **13** 419
[44] Lee S, Min S-J and Eigenmann R 2009 OpenMP to GPGPU *ACM Sigplan Notices* **44** 101–10
[45] The OpenACC application programming interface version 3.3 (available at: https://www.openacc.org/sites/default/files/inline-images/Specification/OpenACC-3.3-final.pdf) (Accessed November 2022)
[46] Maintz S and Wetzstein M 2018 Strategies to accelerate VASP with GPUs using OpenACC *Proc. Cray User Group* (available at: https://cug.org/proceedings/cug2018_proceedings/includes/files/pap153s2-file1.pdf)
[47] Edwards H C, Trott C R and Sunderland D 2014 Kokkos: enabling manycore performance portability through polymorphic memory access patterns *J. Parallel Distrib. Comput.* **74** 3202–16
[48] Beckingsale D, Burmark J, Hornung R, Jones H, Killian W, Kunen A J, Pearce O, Robinson P, Ryujin B S and Scogland T R 2019 RAJA: portable performance for large-scale scientific applications *2019 IEEE/ACM Int. Workshop on Performance, Portability and Productivity in HPC (P3HPC)* (U.S. Department of Energy Office of Scientific and Technical Information) (https://doi.org/10.1109/p3hpc49587.2019.00012)
[49] Peng J *et al* 2022 Human- and machine-centred designs of molecules and materials for sustainability and decarbonization *Nat. Rev. Mater.* **7** 991–1009
[50] Pilania G 2021 Machine learning in materials science: from explainable predictions to autonomous design *Comput. Mater. Sci.* **193** 13
[51] Wilkinson M D *et al* 2016 The FAIR guiding principles for scientific data management and stewardship *Sci. Data* **3** 160018
[52] Trunschke A *et al* 2020 Towards experimental handbooks in catalysis *Top. Catal.* **63** 1683–99
[53] Smith A, Bhat V, Ai Q and Risko C 2022 Challenges in information-mining the materials literature: a case study and perspective *Chem. Mater.* **34** 4821–7
[54] Marshall C P, Schumann J and Trunschke A 2023 Achieving digital catalysis: strategies for data acquisition, storage and use *Angew. Chem., Int. Ed.* **62** e202302971
[55] Foppa L *et al* 2023 Data-centric heterogeneous catalysis: identifying rules and materials genes of alkane selective oxidation? *J. Am. Chem. Soc.* **145** 3427–42
[56] Trunschke A 2022 Prospects and challenges for autonomous catalyst discovery viewed from an experimental perspective *Catal. Sci. Technol.* **12** 3650–69
[57] Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K and Scheffler M 2009 Ab initio molecular simulations with numeric atom-centered orbitals *Comput. Phys. Commun.* **180** 2175–96
[58] Lu H *et al* 2023 Electronic impurity doping of a 2D hybrid lead iodide perovskite by Bi and Sn *PRX Energy* **2** 023010
[59] Ihrig A C, Wieferink J, Zhang I Y, Ropo M, Ren X, Rinke P, Scheffler M and Blum V 2015 Accurate localized resolution of identity approach for linear-scaling hybrid density functionals and for many-body perturbation theory *New J. Phys.* **17** 093020
[60] Levchenko S V, Ren X, Wieferink J, Johanni R, Rinke P, Blum V and Scheffler M 2015 Hybrid functionals for large periodic systems in an all-electron, numeric atom-centered basis framework *Comput. Phys. Commun.* **192** 60–69
[61] Knuth F, Carbogno C, Atalla V, Blum V and Scheffler M 2015 All-electron formalism for total energy strain derivatives and stress tensor components for numeric atom-centered orbitals *Comput. Phys. Commun.* **190** 33–50
[62] Huhn W P and Blum V 2017 One-hundred-three compound band-structure benchmark of post-self-consistent spin-orbit coupling treatments in density functional theory *Phys. Rev. Mater.* **1** 033803
[63] Marek A, Blum V, Johanni R, Havu V, Lang B, Auckenthaler T, Heinecke A, Bungartz H J and Lederer H 2014 The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science *J. Phys.: Condens. Matter* **26** 213201
[64] Kůs P, Marek A, Köcher S S, Kowalski -H-H, Carbogno C, Scheurer C, Reuter K, Scheffler M and Lederer H 2019 Optimizations of the eigensolvers in the ELPA library *Parallel Comput.* **85** 167–77

[65] Yu V W-Z, Moussa J, Kus P, Marek A, Messmer P, Yoon M, Lederer H and Blum V 2021 GPU-acceleration of the ELPA2 distributed eigensolver for dense symmetric and Hermitian eigenproblems *Comput. Phys. Commun.* **262** 107808
[66] Draxl C and Scheffler M 2018 NOMAD: the FAIR concept for big data-driven materials science *MRS Bull.* **43** 676–82
[67] Scheidgen M *et al* 2023 NOMAD: a distributed web-based platform for managing materials science research data *J. Open Source Softw.* **8** 5388
[68] Ghiringhelli L M *et al* 2023 Shared metadata for data-centric materials science *Sci. Data* **10** 626
[69] Sbailò L, Fekete Á, Ghiringhelli L M and Scheffler M 2022 The NOMAD artificial-intelligence toolkit: turning materials-science data into knowledge and understanding *npj Comput. Mater.* **8** 250
[70] Ragan-Kelley B *et al* 2018 Binder 2.0-reproducible, interactive, sharable environments for science at scale *Proc. 17th Python in Science Conf.* ed F Akici, D Lippa, D Niederhut and M Pacer pp 113–20
[71] Google Research, Google Colaboratory 2018 (available at: https://colab.research.google.com/) (Accessed 19 December 2023)
[72] Janssen J, Surendralal S, Lysogorskiy Y, Todorova M, Hickel T, Drautz R and Neugebauer J 2019 pyiron: an integrated development environment for computational materials science *Comput. Mater. Sci.* **163** 24–36
[73] Yakutovich A V *et al* 2021 AiiDAlab–an ecosystem for developing, executing, and sharing scientific workflows *Comput. Mater. Sci.* **188** 110165
[74] Dunn A, Wang Q, Ganose A, Dopp D and Jain A 2020 Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm *npj Comput. Mater.* **6** 1–10
[75] Barker M *et al* 2022 Introducing the FAIR principles for research software *Sci. Data* **9** 622
[76] Alzubaidi L *et al* 2023 A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications *J. Big Data* **10** 46
[77] Giri S K, Saalmann U and Rost J M 2020 Purifying electron spectra from noisy pulses with machine learning using synthetic Hamilton matrices *Phys. Rev. Lett.* **124** 113201
[78] Giri S K, Alonso L, Saalmann U and Rost J M 2021 Perspectives for analyzing non-linear photoionization spectra with deep neural networks trained with synthetic Hamilton matrices *Farad. Discuss.* **228** 502
[79] Cheung H L, Uvdal P and Mirkhalaf M 2023 Augmentation of scarce data—a new approach for deep-learning modeling of composites (arXiv:2311.14557)
[80] Ghane E, Fagerström M and Mirkhalaf M 2023 Recurrent neural networks and transfer learning for elasto-plasticity in woven composites (arXiv:2311.13434v2)
[81] Giri S K, Saalmann U and Rost J M 2024 in preparation
[82] Selstø S 2022 Absorbers as detectors for unbound quantum systems *Phys. Rev.* A **106** 042213
[83] Leo J, Ge E and Li S 2023 Wasserstein distance in deep learning *SSRN Electron. J.* (https://doi.org/10.2139/ssrn.4368733)
[84] Cressie N and Wikle C K 2011 *Statistics for Spatio-Temporal Data* 1st edn (Wiley)
[85] Ostoja-Starzewski M 2007 *Microstructural Randomness and Scaling in Mechanics of Materials* (Chapman and Hall/CRC) (https://doi.org/10.1201/9781420010275)
[86] Saunders R N, Teferra K, Elwany A, Michopoulos J G and Lagoudas D 2023 Metal AM process-structure-property relational linkages using Gaussian process surrogates *Addit. Manuf.* **62** 103398
[87] Xian R P *et al* 2023 A machine learning route between band mapping and band structure *Nat. Comput. Sci.* **3** 101–14
[88] Kusne A G *et al* 2020 On-the-fly closed-loop materials discovery via Bayesian active learning *Nat. Commun.* **11** 5966
[89] Chen R T Q, Amos B and Nickel M 2020 Neural spatio-temporal point processes *Int. Conf. on Learning Representations* (available at: https://openreview.net/forum?id=XQQA6-So14)
[90] Smith J T H, De Mello S, Kautz J, Linderman S W and Byeon W 2023 Convolutional state space models for long-range spatiotemporal modeling (arXiv:2310.19694)
[91] Chang Z, Koulieris G A and Shum H P H 2023 On the design fundamentals of diffusion models: a survey (arXiv:2306.04542)
[92] Abolhasani M and Kumacheva E 2023 The rise of self-driving labs in chemical and materials sciences *Nat. Synth.* **2** 6

[93] Doi M 2015 *Soft Matter Physics* Reprinted with Correction (Oxford University Press)
[94] Menichetti R, Kanekal K H and Bereau T 2019 Drug–membrane permeability across chemical space *ACS Cent. Sci.* **5** 290–8
[95] Greco C, Melnyk A, Kremer K, Andrienko D and Daoulas K C 2019 Generic model for lamellar self-assembly in conjugated polymers: linking mesoscopic morphology and charge transport in P3HT *Macromolecules* **52** 968–81
[96] Potestio R, Peter C and Kremer K 2014 Computer simulations of soft matter: linking the scales *Entropy* **16** 4199–245
[97] Schmid F 2023 Understanding and modeling polymers: the challenge of multiple scales *ACS Polym. Au* **3** 28–58
[98] Jackson N E, Webb M A and De Pablo J J 2019 Recent advances in machine learning towards multiscale soft materials design *Curr. Opin. Chem. Eng.* **23** 106–14
[99] Ni B and Buehler M J 2024 MechAgents: large language model multi-agent collaborations can solve mechanics problems, generate new data, and integrate knowledge *Extreme Mech. Lett.* **67** 102131
[100] Suryanarayana C 1999 *Non-Equilibrium Processing of Materials* (Elsevier)
[101] Musil F, Grisafi A, Bartók A P, Ortner C, Csányi G and Ceriotti M 2021 Physics-inspired structural representations for molecules and materials *Chem. Rev.* **121** 9759–815
[102] Weinreich J, Lemm D, Von Rudorff G F and Von Lilienfeld O A 2022 Ab initio machine learning of phase space averages *J. Chem. Phys.* **157** 024303
[103] Mohr B, Van Der Mast D and Bereau T 2023 Condensed-phase molecular representation to link structure and thermodynamics in molecular dynamics *J. Chem. Theory. Comput.* **19** 4770–9
[104] Wang J, Olsson S, Wehmeyer C, Pérez A, Charron N E, de Fabritiis G, Noé F and Clementi C 2019 Machine learning of coarse-grained molecular dynamics force fields *ACS Cent. Sci.* **5** 755–67
[105] Durumeric A E P and Voth G A 2019 Adversarial-residual-coarse-graining: applying machine learning theory to systematic molecular coarse-graining *J. Chem. Phys.* **151** 124110
[106] Roters F *et al* 2019 DAMASK—the Düsseldorf Advanced Material Simulation Kit for modeling multi-physics crystal plasticity, thermal, and damage phenomena from the single crystal up to the component scale *Comput. Mater. Sci.* **158** 420–78
[107] Raabe D, Mianroodi J R and Neugebauer J 2023 Computational design of compositionally complex materials *Nat. Comput. Sci.* **3** 198–209
[108] Wu X 1991 Neural network-based material modeling *PhD Thesis* Department of Civil and Environmental Engineering at the University of Illinois Urbana–Champaign, Urbana, Illinois
[109] Yang Z, Yu C-H and Buehler M J 2021 Deep learning model to predict complex stress and strain fields in hierarchical composites *Sci. Adv.* **7** eabd7416
[110] Mianroodi J R, Siboni N H and Raabe D 2021 Teaching solid mechanics to artificial intelligence—a fast solver for heterogeneous materials *npj Comput. Mater.* **7** 99
[111] Khorrami M S, Mianroodi J R, Siboni N H, Goyal P, Svendsen B, Benner P and Raabe D 2023 An artificial neural network for surrogate modeling of stress fields in viscoplastic polycrystalline materials *npj Comput. Mater.* **9** 37
[112] Rashid M M, Pittie T, Chakraborty S and Krishnan N M A 2022 Learning the stress-strain fields in digital composites using Fourier neural operator *iScience* **25** 105452
[113] Ni B and Buehler M J 2024 MechAgents: Large language model multi-agent collaborations can solve mechanics problems, generate new data, and integrate knowledge *Extreme Mech. Lett.* **67** 102131
[114] Roters F, Eisenlohr P, Hantcherli L, Tjahjanto D D, Bieler T R and Raabe D 2010 Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications *Acta Mater.* **58** 1152–211
[115] Wang S, Wang H and Perdikaris P 2021 Learning the solution operator of parametric partial differential equations with physics-informed DeepONets *Sci. Adv.* **7** eabi8605
[116] Li Z, Zheng H, Kovachki N, Jin D, Chen H, Liu B, Azizzadenesheli K 2021 Physics-informed neural operator for learning partial differential equations (arXiv:2111.03794)
[117] Raabe D, Sander B, Friák M, Ma D and Neugebauer J 2007 Theory-guided bottom-up design of β-titanium alloys as biomaterials based on first principles calculations: theory and experiments *Acta Mater.* **55** 4475–87
[118] Sandlöbes S, Friák M, Korte-Kerzel S, Pei Z, Neugebauer J and Raabe D 2017 A rare-earth free magnesium alloy with improved intrinsic ductility *Sci. Rep.* **7** 1–8

[119] Goyal P, Duff I P and Benner P 2023 Guaranteed stable quadratic models and their applications in SINDy and operator inference (arXiv:2308.13819)
[120] Sasidhar K N, Siboni N H, Mianroodi J R, Rohwerder M, Neugebauer J and Raabe D 2022 Deep learning framework for uncovering compositional and environmental contributions to pitting resistance in passivating alloys *npj Mater. Degrad.* **6** 71
[121] Lusch B, Kutz J N and Brunton S L 2018 Deep learning for universal linear embeddings of nonlinear dynamics *Nat. Commun.* **9** 4950
[122] Wang N, C.Freysoldt C, Zhang S, Liebscher C H and Neugebauer J 2021 Segmentation of static and dynamic atomic-resolution microscopy data sets with unsupervised machine learning using local symmetry descriptors *Microsc. Microanal.* **27** 1454–64
[123] Saxena A, Polin N, Kusampudi N, Katnagallu S, Molina-Luna L, Gutfleisch O, Berkels B, Gault B, Neugebauer J and Freysoldt C 2023 A machine learning framework for quantifying chemical segregation and microstructural features in atom probe tomography data *Microsc. Microanal.* **29** 1658–70
[124] Kalinin S V, Dyck O, Jesse S and Ziatdinov M 2021 Exploring order parameters and dynamic processes in disordered systems via variational autoencoders *Sci. Adv.* **7** eabd5084
[125] eLabFTW—a free and open source electronic lab notebook (available at: https://www.elabftw.net/) (Accessed 24 October 2023)
[126] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002-1–11
[127] APT-HDF5 file specification (available at: http://fieldemission.org/files/APT-HDF5-2020-10.pdf) (Accessed 24 October 2023)
[128] Electron microscopy datasets (available at: https://emdatasets.com/format/) (Accessed 24 October 2023)
[129] Hyperspy user guide io module (available at: https://hyperspy.org/hyperspy-doc/current/user_guide/io.html) (Accessed 24 October 2023)
[130] Ott S, Hebenstreit K, Liévin V, Hother C E, Moradi M, Mayrhauser M, Praas R, Winther O and Samwald M 2023 ThoughtSource: a central hub for large language model reasoning data *Sci. Data* **10** 1–12
[131] Jablonka K M *et al* 2023 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon *Digit. Discov.* 2 1233–50
[132] Park Y J, Kaplan D, Ren Z, Hsu C W, Li C, Xu H, Li S and Li J 2023 Can ChatGPT be used to generate scientific hypotheses? *J. Mater.* **10** 1–37
[133] Szymanski N J *et al* 2023 An autonomous laboratory for the accelerated synthesis of novel materials *Nature* **624** 86–91
[134] Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, Persson K A, Ceder G and Jain A 2019 Unsupervised word embeddings capture latent knowledge from materials science literature *Nature* **571** 95–98
[135] Zheng Z, Zhang O, Borgs C, Chayes J T and Yaghi O M 2023 ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis *J. Am. Chem. Soc.* **145** 18048–62
[136] Kim E, Huang K, Saunders A, McCallum A, Ceder G and Olivetti E 2017 Materials synthesis insights from scientific literature via text extraction and machine learning *Chem. Mater.* **29** 9436–44
[137] Gupta T, Zaki M, Krishnan N M A and Mausam M 2022 MatSciBERT: a materials domain language model for text mining and information extraction *npj Comput. Mater.* **8** 1–11
[138] Pei Z, Yin J, Liaw P K and Raabe D 2023 Toward the design of ultrahigh-entropy alloys via mining six million texts *Nat. Commun.* **14** 54
[139] Krenn M and Zeilinger A 2020 Predicting research trends with semantic and neural networks with an application in quantum physics *Proc. Natl Acad. Sci. USA* **117** 1910–6
[140] An Y *et al* 2022 Exploring pre-trained language models to build knowledge graph for metal-organic frameworks (MOFs) *2022 IEEE Int. Conf. on Big Data (Big Data)* (*Osaka, Japan*) pp 3651–8
[141] Devi M A, Prakash C P S, Chinnannavar R P, Joshi V P, Palada R S and Dixit R 2020 An informatic approach to predict the mechanical properties of aluminum alloys using machine learning techniques *2020 Int. Conf. on Smart Electronics and Communication (ICOSEC)* (*Trichy, India*) pp 536–41
[142] Zhao X, Greenberg J, An Y and Hu X T 2021 Fine-tuning BERT model for materials named entity recognition *2021 IEEE Int. Conf. on Big Data (Big Data)* (*Orlando, FL, USA*) pp 3717–20

[143] Sasidhar K N, Siboni N H, Mianroodi J R, Rohwerder M, Neugebauer J and Raabe D 2023 Enhancing corrosion-resistant alloy design through natural language processing and deep learning *Sci. Adv.* **9** 7992
[144] Yin J, Bose A, Cong G, Lyngaas I and Anthony Q 2024 Comparative study of large language model architectures on frontier (arXiv:2402.00691)
[145] Kirklin S, Meredig B and Wolverton C 2013 High-throughput computational screening of new Li-ion battery anode materials *Adv. Energy Mater.* **3** 252–62
[146] Rodríguez-Martínez X, Pascual-San-José E and Campoy-Quiles M 2021 Accelerating organic solar cell material's discovery: high-throughput screening and big data *Energy Environ. Sci.* **14** 3301–22
[147] Bajorath J 2002 Integration of virtual and high-throughput screening *Nat. Rev. Drug Discov.* **1** 882–94
[148] Merchant A, Batzner S, Schoenholz S S, Aykol M, Cheon G and Cubuk E D 2023 Scaling deep learning for materials discovery *Nature* **624** 80–85
[149] Andersen C W *et al* 2021 OPTIMADE, an API for exchanging materials data *Sci. Data* **8** 217
[150] Pyzer-Knapp E O, Suh C, Gómez-Bombarelli R, Aguilera-Iparraguirre J and Aspuru-Guzik A 2015 What is high-throughput virtual screening? A perspective from organic materials discovery *Annu. Rev. Mater. Res.* **45** 195–216
[151] Settles B 2011 From theories to queries: active learning in practice *Active Learning and Experimental Design Workshop in Conjunction with AISTATS 2010* vol 16 (Proceedings of Machine Learning Research) p 1
[152] Li K, Persaud D, Choudhary K, DeCost B, Greenwood M and Hattrick-Simpers J 2023 Exploiting redundancy in large materials datasets for efficient machine learning with less data *Nat. Commun.* **14** 7283
[153] Zhang H, Chen W W, Rondinelli J M and Chen W 2023 ET-AL: entropy-targeted active learning for bias mitigation in materials data *Appl. Phys. Rev.* **10** 021403
[154] Todorović M, Gutmann M U, Corander J and Rinke P 2019 Bayesian inference of atomistic structure in functional materials *npj Comput. Mater.* **5** 35
[155] Curtarolo S *et al* 2012 AFLOW: an automatic framework for high-throughput materials discovery *Comput. Mater. Sci.* **58** 218–26
[156] Mathew K *et al* 2017 Atomate: a high-level interface to generate, execute, and analyze computational materials science workflows *Comput. Mater. Sci.* **139** 140–52
[157] Pizzi G, Cepellotti A, Sabatini R, Marzari N and Kozinsky B 2016 AiiDA: automated interactive infrastructure and database for computational science *Comput. Mater. Sci.* **111** 218–30
[158] Foumani Z Z, Shishehbor M, Yousefpour A and Bostanabad R 2023 Multi-fidelity cost-aware Bayesian optimization *Comput. Methods Appl. Mech. Eng.* **407** 115937
[159] Purcell T A R, Scheffler M, Ghiringhelli L M and Carbogno C 2023 Accelerating materials-space exploration for thermal insulators by mapping materials properties via artificial intelligence *npj Comput. Mater.* **9** 112
[160] Kalinin S V *et al* 2022 Machine learning in scanning transmission electron microscopy *Nat. Rev. Methods Primers* **2** 11
[161] Spurgeon S R *et al* 2021 Towards data-driven next-generation transmission electron microscopy *Nat. Mater.* **20** 274–9
[162] Jesse S, Chi M, Belianinov A, Beekman C, Kalinin S V, Borisevich A Y and Lupini A R 2016 Big data analytics for scanning transmission electron microscopy ptychography *Sci. Rep.* **6** 1–8
[163] Cautaerts N, Crout P, Ånes H W, Prestat E, Jeong J, Dehm G and Liebscher C H 2022 Free, flexible and fast: orientation mapping using the multi-core and GPU-accelerated template matching capabilities in the Python-based open source 4D-STEM analysis toolbox Pyxem *Ultramicroscopy* **237** 113517
[164] Leitherer A, Yeo B C, Liebscher C H and Ghiringhelli L M 2023 Automatic identification of crystal structures and interfaces via artificial-intelligence-based electron microscopy *npj Comput. Mater.* **9** 179
[165] Yin W *et al* 2020 A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy *Nat. Commun.* **11** 1–12
[166] Mukherjee D *et al* 2022 A roadmap for edge computing enabled automated multidimensional transmission electron microscopy *Micros. Today* **30** 10–19
[167] Treder K P, Huang C, Kim J S and Kirkland A I 2022 Applications of deep learning in electron microscopy *Microscopy* **71** i100–15

[168] Madsen J, Liu P, Kling J, Wagner J B, Hansen T W, Winther O and Schiøtz J 2018 A deep learning approach to identify local structures in atomic-resolution transmission electron microscopy images *Adv. Theory Simul.* **1** 1800037
[169] Botifoll M, Pinto-Huguet I and Arbiol J 2022 Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook *Nanoscale Horiz.* **7** 1427–77
[170] Friedrich T, Yu C-P, Verbeeck J and Van Aert S 2023 Phase object reconstruction for 4D-STEM using deep learning *Microsc. Microanal.* **29** 395–407
[171] Wang F, Eljarrat A, Müller J, Henninen T R, Erni R and Koch C T 2020 Multi-resolution convolutional neural networks for inverse problems *Sci. Rep.* **10** 5730
[172] Ziatdinov M, Dyck O, Maksov A, Li X, Sang X, Xiao K, Unocic R R, Vasudevan R, Jesse S and Kalinin S V 2017 Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations *ACS Nano* **11** 12742–52
[173] Munshi J, Rakowski A, Savitzky B H, Zeltmann S E, Ciston J, Henderson M, Cholia S, Minor A M, Chan M K Y and Ophus C 2022 Disentangling multiple scattering with deep learning: application to strain mapping from electron diffraction patterns *npj Comput. Mater.* **8** 254
[174] Bertoni G, Rotunno E, Marsmans D, Tiemeijer P, Tavabi A H, Dunin-Borkowski R E and Grillo V 2023 Near-real-time diagnosis of electron optical phase aberrations in scanning transmission electron microscopy using an artificial neural network *Ultramicroscopy* **245** 113663
[175] Schloz M, Müller J, Pekin T C, Van den Broek W, Madsen J, Susi T and Koch C T 2023 Deep reinforcement learning for data-driven adaptive scanning in ptychography *Sci. Rep.* **13** 8732
[176] Gladyshev A, Schloz M, Pekin T C and Koch C T 2022 Comparison of compression methods for ptychographic reconstructions through decomposition of the diffraction patterns in orthonormal bases *Microsc. Microanal.* **28** 394–7
[177] Gault B, Chiaramonti A, Cojocaru-Mirédin O, Stender P, Dubosq R, Freysoldt C, Makineni S K, Li T, Moody M and Cairney J M 2021 Atom probe tomography *Nat. Rev. Method Primers* **1** 51
[178] Marquis E A and Hyde J M 2010 Applications of atom-probe tomography to the characterisation of solute behaviours *Mater. Sci. Eng.* R **69** 37–62
[179] Haley D, London A J and Moody M P 2020 Processing APT spectral backgrounds for improved quantification *Microsc. Microanal.* **26** 964–77
[180] Meier M S, Bagot P A J, Moody M P and Haley D 2023 Large-scale atom probe tomography data mining: methods and application to inform hydrogen behavior *Microsc. Microanal.* **29** 879–89
[181] Li Y, Zhou X, Colnaghi T, Wei Y, Marek A, Li H, Bauer S, Rampp M and Stephenson L T 2021 Convolutional neural network-assisted recognition of nanoscale $L1_2$ ordered structures in face-centred cubic alloys *npj Comput. Mater.* **7** 1–9
[182] Wei Y *et al* 2021 Machine-learning-enhanced time-of-flight mass spectrometry analysis *Patterns* **2** 100192
[183] Li Y *et al* 2023 Quantitative three-dimensional imaging of chemical short-range order via machine learning enhanced atom probe tomography *Nat. Commun.* **14** 7410
[184] Kühbach M, Bajaj P, Zhao H, Çelik M H M H, Jägle E A and Gault B 2021 On strong-scaling and open-source tools for analyzing atom probe tomography data *npj Comput. Mater.* **7** 1–10
[185] Humphreys J, Lan R and Tao S 2020 Development and recent progress on ammonia synthesis catalysts for Haber–Bosch process *Adv. Energy Sustain. Res.* **2** 2000043
[186] Foster S L, Bakovic S I P, Duda R D, Maheshwari S, Milton R D, Minteer S D, Janik M J, Renner J N and Greenlee L F 2018 Catalysts for nitrogen reduction to ammonia *Nat. Catal.* **1** 490–500
[187] Li H, Jiao Y, Davey K and Qiao S 2023 Data-driven machine learning for understanding surface structures of heterogeneous catalysts *Angew. Chem.* **135** e202216383
[188] Burger B *et al* 2020 A mobile robotic chemist *Nature* **583** 237–41
[189] Mou T, Pillai H S, Wang S, Wan M, Han X, Schweitzer N M, Che F and Xin H 2023 Bridging the complexity gap in computational heterogeneous catalysis with machine learning *Nat. Catal.* **6** 122–36
[190] Margraf J T, Jung H-W, Scheurer C and Reuter K 2023 Exploring catalytic reaction networks with machine learning *Nat. Catal.* **6** 112–21
[191] Taniike T and Takahashi K 2023 The value of negative results in data-driven catalysis research *Nat. Catal.* **6** 108–11

[192] Chanussot L *et al* 2021 Open catalyst 2020 (OC20) dataset and community challenges *ACS Catal.* **11** 6059–72
[193] Tran R *et al* 2023 The open catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts *ACS Catal.* **13** 3066–84
[194] Guinier A and Fournet G 1955 *Small-Angle Scattering of X-Rays* (Wiley)
[195] Fratzl P, Jakob H F, Rinnerthaler S, Roschger P and Klaushofer K 1997 Position-resolved small-angle x-ray scattering of complex biological materials *J. Appl. Crystallogr.* **30** 765–9
[196] Liebi M, Georgiadis M, Menzel A, Schneider P, Kohlbrecher J, Bunk O and Guizar-Sicairos M 2015 Nanostructure surveys of macroscopic specimens by small-angle scattering tensor tomography *Nature* **527** 349
[197] Schaff F, Bech M, Zaslansky P, Jud C, Liebi M, Guizar-Sicairos M and Pfeiffer F 2015 Six-dimensional real and reciprocal space small-angle x-ray scattering tomography *Nature* **527** 353–8
[198] Märten A, Fratzl P, Paris O and Zaslansky P 2010 On the mineral in collagen of human crown dentine *Biomaterials* **31** 5479–90
[199] Benecke G *et al* 2014 A customizable software for fast reduction and analysis of large x-ray scattering data sets: applications of the new DPDAK package to small-angle x-ray scattering and grazing-incidence small-angle x-ray scattering *J. Appl. Crystallogr.* **47** 1797–803
[200] Zhou Z *et al* 2023 A machine learning model for textured x-ray scattering and diffraction image denoising *npj Comput. Mater.* **9** 58
[201] De Falco P *et al* 2021 Tomographic x-ray scattering based on invariant reconstruction: analysis of the 3D nanostructure of bovine bone *J. Appl. Crystallogr.* **54** 486–97
[202] Omori N E, Bobitan A D, Vamvakeros A, Beale A M and Jacques S D M 2023 Recent developments in x-ray diffraction/scattering computed tomography for materials science *Phil. Trans. R. Soc.* A **381** 20220350
[203] Anker A S, Butler K T, Selvan R and Jensen K M O 2023 Machine learning for analysis of experimental scattering and spectroscopy data in materials chemistry *Chem. Sci.* **14** 14003–19
[204] Röding M, Tomaszewski P, Yu S, Borg M and Rönnols J 2022 Machine learning-accelerated small-angle x-ray scattering analysis of disordered two- and three-phase materials *Front. Mater.* **9** 956839
[205] Beaucage P A and Martin T B 2023 The autonomous formulation laboratory: an open liquid handling platform for formulation discovery using x-ray and neutron scattering *Chem. Mater.* **35** 846–52
[206] Yager K G, Majewski P W, Noack M M and Fukuto M 2023 Autonomous x-ray scattering *Nanotechnology* **34** 322001
[207] Ouyang R, Curtarolo S, Ahmetcik E, Scheffler M and Ghiringhelli L M 2018 SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates *Phys. Rev. Mater.* **2** 083802
[208] Purcell T A R, Scheffler M, Carbogno C and Ghiringhelli L M 2022 SISSO++: a C++ implementation of the sure-independence screening and sparsifying operator approach *J. Open Source Softw.* **7** 3960
[209] Cairney J M, Rajan K, Haley D, Gault B, Bagot P A J, Choi P-P, Felfer P J, Ringer S P, Marceau R K W and Moody M P 2015 Mining information from atom probe data *Ultramicroscopy* **159** 324–37