https://doi.org/10.1038/s41578-023-00540-6

Simulations in the era of exascale computing

Choongseok Chang, Volker L. Deringer, Kalpana S. Katti, Veronique Van Speybroeck & Christopher M. Wolverton

Exascale computers — supercomputers that can perform 10^{18} floating point operations per second — started coming online in 2022: in the United States, Frontier launched as the first public exascale supercomputer and Aurora is due to open soon; OceanLight and Tianhe-3 are operational in China; and JUPITER is due to launch in 2023 in Europe. Supercomputers offer unprecedented opportunities for modelling complex materials. In this Viewpoint, five researchers working on different types of materials discuss the most promising directions in computational materials science.

As new supercomputers come online, what advances do you think they will unlock in your area of research?

Veronique Van Speybroeck: More powerful supercomputers will certainly unlock new possibilities. However, the advances in the field of computational chemistry and materials modelling result from both the massive expansion of computing power and the development of innovative algorithms and methods[1]. With the upcoming exascale computing facilities, we can expect simulations of more realistic models of materials, with more atoms and having more realistic length scales, and with the inclusion of defects. It should also become possible to follow materials' behaviour over longer timescales. The exascale computing era might provide an important impetus to further close the gap between experimental observations and modelling. In the past two decades, modelling evolved from an explanatory tool for experimental observations to an essential feature in synergistic modelling–experimental efforts to discover tailor-made materials. Ideally, in the future, the materials discovery chain can be further reduced, and modelling could predict how materials should be modified at the atomic scale to obtain desired functional macroscopic behaviours. This is a very ambitious goal, towards which exascale computing facilities may play an important role. However, many hurdles are yet to be overcome to integrate these very strong computers into the materials modelling ecosystem. Major programming and coding efforts will be necessary to adapt or develop new codes to efficiently unlock the power of these high-performance computers.

Choongseok Chang: As new exascale computers come online, in fusion research we will be able to unlock the critically important heat-load footprint physics for ITER and future magnetic fusion reactors. Simple data explorations based on information from present tokamaks have shown that in ITER the heat-load footprint on the surface of the divertor, which removes waste material produced by the fusion reaction and protects the surrounding walls, could be very narrow. This means that the impinging power density will be above the material heat-load limit unless continuous complex monitoring and operation control are implemented while the fusion process is ongoing. However, because ITER is much bigger and hotter than present tokamaks, there is no guarantee that it will be in the same physics regime. The new exascale computers will be able to study this physics under more realistic conditions and with higher fidelity, making operation of ITER much more reliable.

Kalpana S. Katti: Modelling accurately and efficiently hierarchical and coupled biological systems such as cells, tissues and organs is challenging. Their properties and behaviours are determined by molecular-scale interactions, composition and structural conformations, factors that vary over a wide range of length and temporal scales. Moreover, in most instances, their responses are dictated by coupled chemical, electrical and mechanical properties and by stimuli, and the role of the evolution of the structure with remodelling and ageing, as well as the system's hierarchy, also needs to be incorporated in models. New supercomputers can help build precise models and can run simulations for relatively long periods of time to obtain more meaningful results in physiologically relevant time frames. Such models can often be confirmed by experimental observations.

The ability to build all-atomic models of large macromolecules and hybrid organic–inorganic molecular systems and run the simulations for extended periods to study interactions, chemical reactions and conformational changes, coupled to the possibility of evaluating the mechanical response of large agglomerations of molecules, would allow us to capture 'material' behaviours accurately in multiscale models. The leap in the ability of the new computers to model responses of realistic material systems at the atomic scale would enable the development of large models of biological composites. These models would still capture the atomistic nuances to some extent, using advances in multiscale and coarse-grained modelling. At higher length scales, modelling techniques such as finite-element modelling and discrete element modelling could be used at very small microstructure and at nanostructure-level length scales to capture the microstructural subtleties of biological systems.

Advances in modelling coupled chemical, electrical, thermal and mechanical behaviours are enabling models that closely mimic real systems. I believe the most remarkable advances that will result from new supercomputers, or at least the most impactful, will be in biology, through the unlocking of the understanding of biological function for tissue regeneration, and of diseases such as neurological disorders and cancer progression and metastasis.

Volker L. Deringer: Nowadays, we are seeing a degree of realism in atomic-scale computer simulations that was plainly unthinkable a few years ago. For nanoscale devices, the length scales of experimental characterization (say, by high-resolution electron microscopy) and atomistic simulation have actually converged. We are therefore about to see a step change in how simulations are used: traditionally, they have followed after the fact, to explain an experiment after it had been done. We are moving towards a world where supercomputers guide experiments on structurally complex materials before they are done – not 'just'

finding promising synthesis targets (which is a feat in itself!) but describing the entire formation process on the atomic scale, and predicting its reactants and parameters.

In your field, what are the main challenges for simulations?

KSK: My group works on multiscale modelling of biological systems, especially cancer cells to study tumour progression. Challenges include a computational infrastructure that is inadequate to build atomic-scale models of very large multiprotein and hybrid organic–inorganic systems that model the complexity of biological systems realistically and simulate them efficiently. Also, there is a need for efficient analytical and computational tools to bridge the vast length and temporal scales intrinsic to biological systems. Finally, I believe that in the field of medicine, an area largely fuelled by experiments, one of the most significant challenges is the buy-in of computational results by the medical community: faith in the reliability of computational efforts in biology is growing, but needs to further increase.

VVS: One of the biggest challenges within materials modelling is modelling dynamic processes in realistic functional nanomaterials at operating conditions, yielding information that is comparable to experimental observables[2]. The applications of functional nanomaterials are very broad and include catalysis, separation, capture, and energy storage and conversion. One example of an important material class is zeolites, which have a proven track record for the conversion of traditional petrochemical feedstocks, but also show great potential to convert new non-fossil feedstocks such as biomass or recycle CO_2 into high-value chemicals. However, to optimize their functionality under complex conditions, zeolites need to be meticulously tuned at the molecular level to make them highly selective and active in the desired operation window. Modelling can help tremendously, provided that simulations can capture this complexity.

The complexity originates from various factors. First, it is important to realize that materials are far from perfect: they have defects that dramatically impact their behaviour. Second, the behaviour of materials is very much dependent on the conditions in which they work. Prototypical examples are found in catalysis, where the desired activity and selectivity is only obtained within a very narrow temperature window. When the conditions are slightly changed, materials become less active or even inactive. Catalysts may thus show an on–off behaviour comparable to that of enzymes. Lastly, materials are living objects, with very complicated dynamics covering intrinsic timescales that span many orders of magnitude, from very fast motions associated with localized vibrations to slow global deformation modes.

The main bottleneck of current simulations is the inconsistency between attainable theoretical length scales and timescales and experimentally relevant scales. Theoretically, accessible length scales and timescales are far shorter than the seconds-to-hours timescale and micrometre length scale necessary to simulate transitions in realistic materials. We need methods that can reach the macroscopic scale starting from the atomic level. Having access to substantially more powerful computers will certainly help to further close the length and timescale gap, as many more energy and force evaluations will be possible in shorter times. However, even when having access to massive computer power, the question of how to produce reliable atomic representations for mesoscopic systems remains open.

CSC: In magnetic fusion, the plasma–material interaction is one of the biggest challenges for simulations. To understand and predict the physics of the plasma–material interaction, it is necessary to be able to predict the three-dimensional distribution of density, temperature and flow of the boundary plasma, which is in contact with or near the odd-shaped material wall. Especially important is the highly localized exhaust of the plasma heat that bombards the specially armoured divertor plates along narrow toroidal channels. Unless we either radiate away or spread the exhaust heat, premature erosion of the divertor plate could make magnetic fusion reactors economically non-competitive. The plasma properties in the boundary region are non-locally connected to those of the plasma in the core fusion region via the large-scale kinetic motion of plasma particles and electromagnetic microturbulence in the stationary phase of the energy production operation. It is known experimentally that the local optimization of the plasma near the divertor could harm the fusion performance in the burning core, requiring an integrated optimization. This is a non-local, nonlinear, multiscale kinetic physics problem in a complex geometry that currently constitutes an insurmountable simulation challenge.

VLD: Simulations are becoming more and more abundant, and that brings challenges of its own. One pertains to computational infrastructure: we can no longer afford to spend our time running (many) individual simulations by hand, and instead we need to invest in efficient workflows and automation[3]. Another challenge is validation: making sure that results are physically meaningful, tested against reliable benchmarks and predictive. Validation has always been important, of course, but it gets more difficult to do as the

The contributors

Choongseok (CS) Chang is the head of the multi-institutional multidisciplinary SciDAC Partnership Center for High-fidelity Boundary Plasma Simulation (HBPS), headquartered at the Princeton Plasma Physics Laboratory (PPPL) at Princeton University, awarded by the DOE Office of Science. He is a fellow of the American Physical Society. Before he joined PPPL, he was a physics professor at KAIST in South Korea and, jointly, a tenured research professor at the Courant Institute of Mathematical Sciences at New York University. His speciality is extreme-scale simulations of complex nonlinear multiscale plasma.

Volker L. Deringer is Associate Professor of Theoretical and Computational Inorganic Chemistry at the University of Oxford. His group's research aims at understanding the connections between atomic-scale structure and properties in complex (often, amorphous) materials, driven by emerging machine-learning approaches.

Kalpana S. Katti is a University Distinguished Professor at North Dakota State University and the director of the Center for Engineered Cancer Testbeds. Her primary area of research is in tissue engineering, cancer scaffolds and biomimetics, and her group introduced new engineered nanoclay nanocomposites for biomedical applications.

Veronique Van Speybroeck is full professor at Ghent University in the field of molecular modelling and head of the Center for Molecular Modelling. She has a record of contributions in the field of modelling nanoporous materials for catalysis and adsorption, such as zeolites, metal–organic frameworks and covalent organic frameworks, and she developed methods to determine first-principle kinetics and molecular dynamics simulations of complex chemical transformations in nanoporous materials. Recently, she has been exploring methods to also model spatially extended nanostructured materials in her endeavour to model materials and processes as realistically as possible.

Christopher M. Wolverton is the Jerome B. Cohen Professor of Materials Science and Engineering at Northwestern University. His group's research is centred on computational materials science, and specifically first-principles quantum-mechanical simulation tools, and on the use of machine-learning tools to explore materials datasets and discover new materials.

volume of simulation data increases, and as machine-learned (data-driven, rather than physics-based) models become abundant.

Closely connected to the first point, we need to improve on the open availability of data and code, and we need to further encourage colleagues to share. Behind many of the structure pictures you see in publications there are gigabytes of data, and having access to them can be useful in multiple ways (sometimes in ways the original author hasn't thought of). Open availability of data is not without challenges: as a community, we need to find new ways to acknowledge the creation, curation and sharing of large datasets and to reward best practice. And we need to ensure that there is a certain type of standardization in what is shared and how, which is currently largely missing.

What would you say is the most promising recent development in your field?

Christopher M. Wolverton: In the field of first-principles electronic-structure calculations, density functional theory (DFT) has become a ubiquitous, powerful tool to address many problems in materials science, condensed matter physics and solid-state chemistry. Two prominent developments over the past decade are the use of high-throughput DFT, and the related topic of the use of machine-learning (ML) and artificial intelligence (AI) approaches trained on DFT-computed properties.

High-throughput DFT calculations rely on the fact that many types of calculations can be performed relatively routinely these days, and the tasks of setting up the calculations, performing them, correcting for errors and storing results can all be automated. These automated workflows have led to the development of several large-scale databases of DFT-computed properties of materials, and researchers from around the world now utilize these databases as part of their research efforts. Many of the calculations that underlie these databases were developed years or even decades ago, and so one might wonder whether the fact that so much of the community is doing 'many calculations' rather than new types of calculations indicate that we have collectively run out of ideas. I, of course, don't believe that is the case, but rather that the use of these high-throughput calculations and the resulting large datasets have allowed us to ask questions that we never thought to ask before.

For instance, large materials datasets have led quite naturally to the advent of ML and AI approaches to predict materials properties. These approaches essentially rely on training a machine to learn the relationships between material composition, structure and properties, by using hundreds or thousands (or more) of examples. The field has been active for about a decade now, so it is still fairly new, but extremely active. So, the future of this field is likely to produce exciting new directions, some of which I suspect we cannot even envision at the moment. Interatomic potentials have also experienced a renaissance in recent years. The use of large DFT datasets to train machine-learned interatomic potentials has resulted in unprecedented accuracy and transferability of these approaches. These approaches are much less computationally expensive than DFT, and hence offer the possibility to explore systems that are currently out of the reach of direct DFT.

VLD: There is a surge of interest in ML methods at the moment: the science and craft of building computational models that 'learn' from very large datasets. There is a lot of buzz, but it is backed up by very exciting possibilities indeed.

In my field of inorganic materials modelling, we have come a long way in building models of potential-energy surfaces, referred to as ML-based 'interatomic potentials' or 'force fields'. If trained and validated correctly, they make it possible to run quantum-accurate simulations at a small fraction of the computational cost (compared with direct quantum-mechanical computations), and therefore they are emerging tools for real-world materials modelling[4]. ML potentials predict energies and forces on atoms; beyond that, there is scope to integrate them with electronic-structure theory more widely[5,6], bringing together the (previously rather distinct) questions of where the atoms and where the electrons are. Over the past few years, a range of ML potential fitting methods has become widely available, often with open-source code, and they have been interfaced with mainstream simulation software for large-scale 'production' simulations. New students in the field can now get started with ML-driven materials modelling within a matter of weeks!

VVS: DFT revolutionized the field of chemistry and materials modelling in the late twentieth century by enabling quantum-mechanical simulations of systems of increasing size with much better accuracy than mean-field wavefunction-based methods. However, DFT has reached its limits in terms of both accuracy and number of atoms it can simulate. Currently, major breakthroughs are being obtained by integrating ML techniques within the field of chemistry and materials science[7]. With ML potentials, one generates a numerical potential by nonlinear regression techniques based on underlying DFT data[8]. This allows one to scale up the size of the systems by orders of magnitude. Thus the integration of ML, and in particular deep-learning techniques, is an important element to further close the gap between what can be simulated at the molecular scale and macroscopic functional behaviour.

Although major methodological advances have been made in the field of AI, their integration into the simulation of realistic nanostructured materials with substantial complexity poses enormous challenges, and a close cooperation between various fields will be necessary. The new exascale computing systems raise the question of how to design an efficient heterogeneous computing approach with an optimal use of central processing units (CPUs) and graphical processing units (GPUs) for dedicated tasks in the materials modelling chain. These new approaches will generate enormous amounts of data, whose management and efficient exploitation in advanced AI models will pose major challenges.

KSK: The increased efficiency of modelling codes and the improvements in massively parallel computing infrastructure, including more efficient CPUs, have been important advances. Together with the migration of codes to exploit the power of GPUs, they have led to larger models and more accurate and efficient simulations. The availability of visualization engines and the improved quality and efficiency of pre- and post-processing software have helped interpret simulation results more quickly and enabled the discovery of important cellular and molecular mechanisms.

CSC: Recent first-principles simulations performed on the current largest US supercomputers (Summit at Oak Ridge National Laboratory, Theta at Argonne National Laboratory and Polaris at the National Energy Research Scientific Computing Center) suggest that the ITER boundary plasma will be in a different physics regime from that of present tokamaks, suffering from different types of turbulence that will spread the heat load over a wider area on the material surface[9,10]. This result raises the hope that ITER will be able to produce tenfold more energy than it will receive as input without prematurely damaging the divertor.

What do you think is an intractable problem now that you hope will be solved within the next 5–10 years?

KSK: The computational modelling of diseases will lead to transformational discoveries and therapies for diseases that are currently considered incurable, such as metastatic cancer, Alzheimer disease and autoimmune diseases. Important advancements have been made in characterization techniques, advanced animal models, early detection and therapies to address the symptoms of these health conditions. Yet the fundamental mechanisms underlying these diseases, some of which manifest at the atomic, nano- and microscale, are currently unknown or inadequately known, and need to be understood to develop cures.

The in silico investigations of diseases will be central to enabling discoveries that will lead to therapies and to eradicating terminal diseases. The investigation of these diseases will require high-resolution multiscale models of whole organs and tissues, from the atomic scale to the macroscale, which will serve as computational 'testbeds'. The next generation of supercomputers should be able to achieve this goal in the coming decade. The trial-and-error methods practised in medicine for many conditions will then be applied to computational models rather than patients. Advances in ML and machine vision should enable computational efficiencies in the development and validation of models. For example, they would accelerate the development of accurate force fields for atomistic simulations, which would narrow down the potential protein–protein and protein–mineral interactions to take into consideration and enable the identification of potential drug molecules to test with the computational testbeds, while ensuring accurate modelling of cellular and tissue morphology, cell migration, tissue formation and tumour development.

CSC: In the next 5–10 years, I envision that thanks to exascale and post-exascale computers the plasma–material interaction physics will be understood, with all the most important multiple phenomena simulated together, including the alpha particles and their helium ash particles (generated from the fusion burn), material-sputtered impurity particles and their migration in the plasma, microturbulence physics, large-scale fluid-type motions and instabilities, and the behaviour of materials at the microscale.

VLD: As was mentioned by my colleagues above, a currently insurmountable problem is fully connecting the atomic and macroscopic scales in simulation. We will continue to make atomic-scale modelling much faster, but even with the fastest ML tools we are unlikely to reach the length scale of centimetres and the timescale of seconds and hours on which 'real' experiments often occur. These are two (currently) intractable problems. Regarding length scale, in the future, we might combine atomistic simulations with larger-scale approaches from materials science and engineering, aiming to develop unified models with adaptive resolution — fine-grained where needed, but only there. Regarding timescale, I expect that there will be a need for using advanced sampling techniques much more routinely, and maybe we will come up with entirely new ideas.

A specific issue with ML models is that the vast majority is trained for specific problems: a new domain of application requires at least the extension and re-training of an existing model, and sometimes the development of a whole new training dataset. Can we construct large 'general chemistry' models that are applicable to multiple scientific questions (and across the periodic table) all at once?

VVS: A very challenging problem is how to deal with materials that cannot be simulated accurately with DFT, such as materials where strong electron correlations are present. This is typically the case for materials that contain transition-metal compounds with complicated spin states or rare-earth compounds with partially filled *f*-electron bands. For those systems, current exchange–correlation functionals fail, and one should resort to much more expensive Green's function or direct wavefunction-based methods, which have very bad scaling behaviour with increasing number of electrons[11]. This intractable scaling behaviour makes it impossible to apply these methods to realistic materials of reasonable size. ML methods may be important to enable further progress, but I also expect fundamentally new elements taken from other fields, like tensor networks inspired from quantum information theory, to play a role in the future. In any case, to progress in this field, various communities will have to cooperate, in particular quantum chemists, many-body physicists and computer scientists.

CMW: The use of DFT to predict new materials has exploded in the past few years. Only a few years ago, the DFT prediction of a novel, stable material warranted publication in a very high-impact journal. Now it's quite routine, and papers are regularly published that predict tens or hundreds of such materials. This leads to the intriguing question: can we actually find all possible stable inorganic materials? This question would have seemed ludicrous a short time ago, but now we can actually debate the idea. What is the total number of stable inorganic compounds, and what fraction of these have we experimentally discovered? How far are we from this 'finish line'? In just the past few years, computational predictions have blossomed, and there are DFT databases that have several times more predicted stable compounds than the total number of experimentally known ones.

Synthesizing even a fraction of these materials and achieving rationally designed, computationally guided synthesis will be a great challenge in the coming years. Pushing the science of synthesis forward is currently an active area of research, and it seems that future developments will push towards predictive, synthetic 'recipes' for producing novel materials. Another grand challenge involves autonomous materials discovery and development via a combination of human-out-of-the-loop experiment, computation and ML/AI. Computational methods to automate experimental characterization, accelerate the prediction of new materials and train AI approaches will play a key role in the development of this nascent field.

More advanced property prediction is also advancing quite quickly. Calculation of properties that are beyond simple DFT total energy calculations are already forming valuable datasets and are likely to form the basis of future datasets. For instance, my group is actively working in constructing high-throughput datasets that move beyond the confines of $T = 0$ K energetics by calculating phonon properties, both in harmonic and anharmonic forms. Harmonic enables the computation of (vibrational) free energies, and anharmonic of phonon-scattering processes and higher-order effects, such as thermal conductivity and temperature-dependent phonon renormalization.

Another means to improve these databases is to incorporate more accurate exchange-correlation functionals (such as meta generalized gradient approximation functionals) in the computational workflows. One also can imagine that AI-based approaches might learn new functionals that are physically more accurate than our best theoretical constructs. Work has already begun in this direction, but the next 5 years will be telling. Have we reached the era of AI-informed DFT or, more generally, AI-informed materials science?

Finally, all these data-driven advances will happen against the backdrop of the use of computation in high-fidelity, physics- and chemistry-informed understanding of materials and processes. These kinds of uses of computation have been present since the advent of these techniques, and have produced innumerable advances in understanding. This work will also progress alongside the data-driven work, and will probably produce new insights and discoveries and generate new understanding that is difficult to forecast.

Is there anything that can be simulated now that you wouldn't have thought possible when you started your career?

CSC: When I started my career in the 1970s, we were able to utilize only simple, localized, linear plasma models. With the subsequent tremendous enhancement in computing power, we can now simulate self-organized, nonlinear and scale-inseparable multiple phenomena to model the interaction of the plasma with the wall in realistic boundary geometries. This was not thought possible at the start of my career. As a result, materials scientists are able to develop their models starting from more realistic plasma particles that enter the material surface with more realistic and dynamical energy and angle.

VVS: Compared with the late 1990s, when I started as a PhD student, enormous progress has been made in terms of system size, accuracy and dynamics in materials simulations. Nowadays, thanks to ingenious algorithms and advanced software and hardware architectures, we are, for example, able to simulate with classical force fields the phase transformations of flexible materials using material model representations with a million atoms[12]. Or we can simulate on-the-fly a catalytic reaction in the pores of a zeolite in the presence of moisture or water, using quantum-mechanical methods[13]. These examples prove that enormous progress can be made thanks to extensive computer power, massive parallelization, usage of GPUs, dedicated software and hardware architectures, and most of all innovative methods and algorithms.

KSK: The construction of all-atom models of whole viruses was a milestone in biology and medicine, and computational immunology helped greatly accelerate the development of a vaccine for COVID-19. However, although these are outstanding achievements compared with the state of the field two decades ago, when my scientific journey began, I expected significantly more advances in computational modelling, including computational modelling of whole organs. The advances in AI have been even more disappointing, with limited success in addressing major diseases that cause untimely deaths and suffering. However, considering the current surge in computational tools, the scientific knowledge we can accumulate thanks to high-precision characterization tools, and the likelihood of buy-in for computational modelling and simulations from medical practitioners and the public, I believe computations are poised to accelerate breakthroughs in the fight against major diseases.

VLD: When I was a doctoral student about a decade ago, I was seriously impressed by seeing simulations of a few hundred atoms at a time. We're able to look at millions now on relatively modest computing systems, and colleagues have shown proof-of-concepts for billions. We have always accepted, tacitly or explicitly, that quantum-mechanically based simulations create highly simplified models of reality (say, a supercell of a crystal with a single defect). Today's and tomorrow's ML-driven simulations are now on the verge of describing real materials in their entirety, all while keeping the quantum-mechanical accuracy: from single defects to nanocrystalline grains; from well-defined crystal surface reconstructions to rough amorphous surfaces or to interfaces in a device; from simplified chemical compositions to the exact homogeneous and heterogeneous mixtures of materials that our colleagues create in the lab. That's nothing short of a qualitative change.

Choongseok Chang[1] ✉, **Volker L. Deringer**[2] ✉, **Kalpana S. Katti**[3] ✉, **Veronique Van Speybroeck**[4] ✉ & **Christopher M. Wolverton**[5] ✉

[1]Princeton Plasma Physics Laboratory, Princeton, NJ, USA. [2]Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford, UK. [3]Department of Civil, Construction and Environmental Engineering, Center for Engineered Cancer Testbeds, Materials and Nanotechnology Program, North Dakota State University, Fargo, ND, USA. [4]Center for Molecular Modelling, Ghent University, Zwijnaarde, Belgium. [5]Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA.

✉e-mail: cschang@pppl.gov; volker.deringer@chem.ox.ac.uk; kalpana.katti@ndsu.edu; veronique.vanspeybroeck@ugent.be; c-wolverton@northwestern.edu

Published online: 13 March 2023

References

1. Ciccotti, G., Dellago, C., Ferrario, M., Hernández, E. R. & Tuckerman, M. E. Molecular simulations: past, present, and future (a Topical Issue in EPJB). *Eur. Phys. J. B* **95**, 3 (2022).
2. Van Speybroeck, V., Vandenhaute, S., Hoffman, A. E. J. & Rogge, S. M. J. Towards modeling spatiotemporal processes in metal–organic frameworks. *Trends Chem.* **3**, 605–619 (2021).
3. George, J. Automation in DFT-based computational materials science. *Trends Chem.* **3**, 697–699 (2021).
4. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
5. Westermayr, J., Gastegger, M., Schütt, K. T. & Maurer, R. J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **154**, 230903 (2021).
6. Ceriotti, M. Beyond potentials: integrated machine learning models for materials. *MRS Bull.* **47**, 1045–1053 (2022).
7. Friederich, P., Hase, F., Proppe, J. & Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **20**, 750–761 (2021).
8. Behler, J. & Csányi, G. Machine learning potentials for extended systems: a perspective. *Eur. Phys. J. B* **94**, 142 (2021).
9. Chang, C. S. et al. Constructing a new predictive scaling formula for ITER's divertor heat-load width informed by a simulation-anchored machine learning. *Phys. Plasmas* **28**, 022501 (2021).
10. Chang, C. S. et al. Gyrokinetic projection of the divertor heat-flux width from present tokamaks to ITER. *Nucl. Fusion* **57**, 116023 (2017).
11. Cui, Z.-H., Zhu, T. & Chan, G. K.-L. Efficient implementation of ab initio quantum embedding in periodic systems: density matrix embedding theory. *J. Chem. Theory Comput.* **16**, 119–129 (2020).
12. Vandenhaute, S., Rogge, S. M. J. & Van Speybroeck, V. Large-scale molecular dynamics simulations reveal new insights into the phase transition mechanisms in MIL-53(Al). *Front. Chem.* **9**, 718920 (2021).
13. Bocus, M., Vanduyfhuys, L., De Proft, F., Weckhuysen, B. M. & Van Speybroeck, V. Mechanistic characterization of zeolite-catalyzed aromatic electrophilic substitution at realistic operating conditions. *JACS Au* **2**, 502–514 (2022).

Competing interests

The authors declare no competing interests.