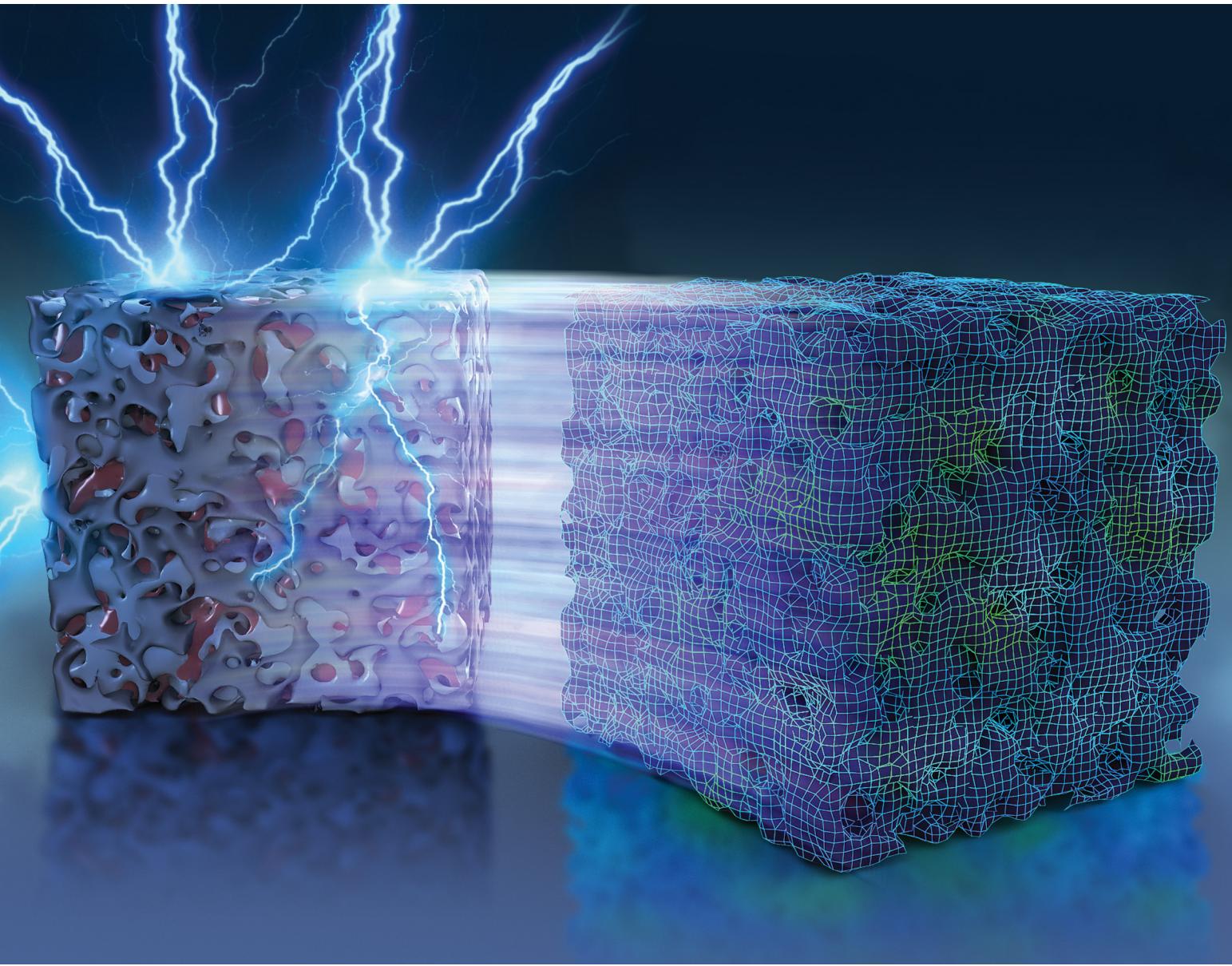


Energy & Environmental Science

rsc.li/ees



ISSN 1754-5706



Cite this: *Energy Environ. Sci.*, 2021, 14, 2549

Towards the digitalisation of porous energy materials: evolution of digital approaches for microstructural design

Zhiqiang Niu, ^{ab} Valerie J. Pinfield, ^a Billy Wu,^c Huizhi Wang,^d Kui Jiao, *^e Dennis Y. C. Leung *^b and Jin Xuan *^a

Porous energy materials are essential components of many energy devices and systems, the development of which have been long plagued by two main challenges. The first is the 'curse of dimensionality', *i.e.* the complex structure–property relationships of energy materials are largely determined by a high-dimensional parameter space. The second challenge is the low efficiency of optimisation/discovery techniques for new energy materials. Digitalisation of porous energy materials is currently being considered as one of the most promising solutions to tackle these issues by transforming all material information into the digital space using reconstruction and imaging data and fusing this with various computational methods. With the help of material digitalisation, the rapid characterisation, the prediction of properties, and the autonomous optimisation of new energy materials can be achieved by using advanced mathematical algorithms. In this paper, we review the evolution of these computational and digital approaches and their typical applications in studying various porous energy materials and devices. Particularly, we address the recent progress of artificial intelligence (AI) in porous energy materials and highlight the successful application of several deep learning methods in microstructural reconstruction and generation, property prediction, and the performance optimisation of energy materials in service. We also provide a perspective on the potential of deep learning methods in achieving autonomous optimisation and discovery of new porous energy materials based on advanced computational modelling and AI techniques.

Received 6th February 2021,
Accepted 1st April 2021

DOI: 10.1039/d1ee00398d

rsc.li/ees

Broader context

Porous energy materials are essential to many energy devices and systems. However, their development has been plagued by two challenges for long periods. The first is the 'curse of dimensionality', *i.e.* the structure–property relationships are largely determined by a high-dimensional parameter space; the second is low efficiency of optimisation or discovery of new energy materials from big sets of candidates. Digitalisation of porous energy materials has been regarded one of the most promising solutions to tackle above issues because it shapes all materials information into digital data space with various advanced computational approaches and imaging techniques. This paper reviews the evolution of computational and digital approaches in different periods. Particularly, we addressed the recent progress of artificial intelligence in microstructures generation, properties prediction, structures design and device performance optimisation. We also proposed a perspective on how to achieve autonomous optimisation and discovery of new porous energy materials based on AI techniques.

1. Introduction

Porous materials are used in a range of energy conversion and storage systems including electrochemical devices,¹ high-capacity gas storage systems,² and carbon capture and sequestration systems,³ to name but a few. The diversity of energy systems and applications results in a broad range of porous energy materials from naturally formed materials such as coal to artificially synthesised structures such as metal foams and carbon-based electrodes. These porous energy materials have numerous physical and chemical properties covering a range of

^a Department of Chemical Engineering, Loughborough University, Loughborough, UK. E-mail: J.Xuan@lboro.ac.uk; Fax: +44 (0)1509 227186;

Tel: +44 (0)1509 227186

^b Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China. E-mail: ycleung@hku.hk

^c Dyson School of Design Engineering, Imperial College London, London, UK

^d Department of Mechanical Engineering, Imperial College London, London, UK

^e State Key Laboratory of Engines, Tianjin University, 135 Yaguan Road, Tianjin, 300350, China. E-mail: kjiao@tju.edu.cn



length scales which determine their suitability for different applications.^{3–6} Thus, understanding the nature of these properties is critical for the development of future energy devices.

However, the microstructures of porous energy materials are complex due to their complicated formation mechanisms and fabrication processes. For example, the formation of the complex morphology found in Li-ion battery electrodes involves mixing active materials, drying to remove solvent and mechanical calendaring,⁷ while electrodes of solid oxide fuel cells (SOFCs) are usually prepared through the creation of an ink consisting of yttria-stabilised zirconia (YSZ) powder-doped with Ni and lanthanum manganite (LSM), which is screen-printed and sintered at temperatures of ~ 1300 °C.⁸ These multi-step synthesis protocols (mechanical, chemical, thermal, *etc.*) subsequently lead to multi-scale characteristics, with features such as pore size or characteristic length ranging widely from the nanometre to millimetre scale. In addition, porous energy materials often involve multi-physical processes when they are in service. For instance, metal foams act with two functions; as mechanical supports and flow distributors in fuel cells.⁹ Porous silicon materials offer reaction sites in photovoltaic systems but also transfer free charge carriers.¹⁰ Simultaneous ionic, electronic and gas transport occur at triple-phase boundaries in the catalyst layers (CL) of fuel cells.¹¹ Thus, accurately describing the complex relationships between structure, properties and performance of these porous energy materials has been challenging due to the multi-scale and multi-physical nature of their performance.

One of the most promising approaches to understand the complicated physics of porous energy materials is to fully digitalise energy materials. The digitalisation of energy materials will transform all material information including their structures, properties and performance into the data space, allowing for *in silico* design. This not only allows in-depth, massive and efficient data analysis to extract meaningful features such as material properties and lifetime, but also to deliver insightful information for design, optimisation and discovery of innovative energy materials. In recent years, the progress of energy material digitalisation has been greatly promoted by increasing effort in developing experimental and computational methods.^{12–17}

Several ambitious initiatives have been announced to promote the exploration of innovative materials by international communities, *e.g.* the Materials Genome Initiative (MGI)¹⁸ in the United States in 2011 and the Material Digitalization Platform¹⁹ in Germany in 2019 which aim to digitalise materials from the atomic level to the system level to accelerate discovery of new materials.^{20–22} In this review, we focus on the emerging development and utilisation of digital technologies towards new understanding of the fundamentals, such as microstructures of porous energy materials, acceleration of the material optimisation and discovery, and improvement of the performance of porous energy materials in various applications. This bridges the fundamental molecular-scale design of materials with their eventual performance in macroscopic energy devices and systems. First, we briefly review the research history of porous energy materials that includes empirical generalisations, basic theories and imaging techniques. Then, the recent progress of powerful

imaging and modelling techniques in modern porous energy materials research are discussed. Finally, we address the emerging trends of artificial intelligence (AI) in porous energy materials and highlight the successful applications of several deep learning methods in microstructural reconstruction, property prediction, and performance optimisation of energy materials. We also provide a perspective on the potential of these deep learning methods in achieving autonomous optimisation and discovery of new porous energy materials based on powerful computational modelling and AI techniques.

2. A brief history of energy materials development: the foundations of empirical, theoretical and imaging techniques

Before the first industrial revolution, naturally occurring materials were widely used (such as wood or charcoal) serving as energy sources for heating food and material-forming (metal-working, earthenware *etc.*). The efficient utilisation of these porous energy materials mainly depended on empirical observations, with few tools or theories available for understanding these energy materials.

Since the first industrial revolution in the 18th century, several porous energy materials for energy supply and storage, such as coal and oil sandstones, have become the most relevant because much of the energy exploitation and utilisation, directly or indirectly, involved them. Both of these materials possess complex microstructures; however, few imaging techniques were available to reveal the full complexity of their microstructures due to the limited resolution of optical microscopes. Even today, the highest resolution of conventional optical microscopes is limited to ~ 0.2 μm due to the diffraction limit, accuracy of lenses or mirrors and/or detector array resolution,²³ so features at length scales smaller than this are not identified with this technique. Fig. 1 shows 2D images of one typical coal and oil sandstone obtained by a low-resolution optical microscope.^{24,25} It is seen that only large-scale features can be distinguished with small-scale structures being unresolved. In addition, these 2D images are thin sections extracted from samples and are therefore insufficient to provide the 3D spatially resolved structures that are needed to fully understand the structural characteristics of porous energy materials.

It has been challenging to predict physical properties of porous energy materials from optical images of their structure. A series of analytical theories were developed for a number of physical processes, such as thermal conduction (Fourier's law, 1822), fluid dynamics (Navier–Stokes equations, 1845), mechanics of materials (Saint-Venant's principle, 1855) and so on. However, these theories were only to describe simple configurations and were not able to characterise the highly complex structures observed in practice until computers were introduced for numerical simulations.²⁶ Thus, empirical or semi-empirical correlations were often preferred in order to relate bulk geometrical characteristics to



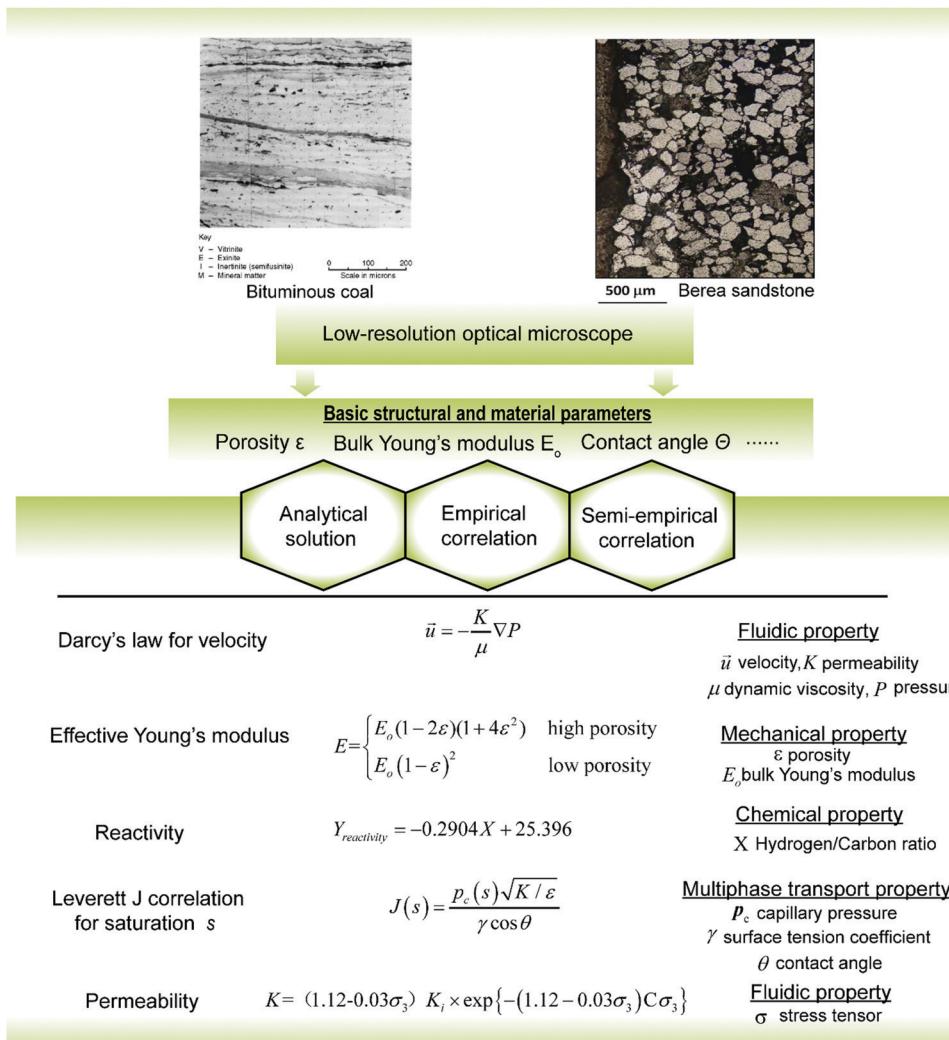


Fig. 1 The content of traditional research of porous energy materials (the images are reproduced from ref. 24 and 25 with permission from Elsevier and the equations are from ref. 27–30).

properties.^{27–29} Several classical correlations for estimating different physical properties are shown in Fig. 1. Although these equations require some assumptions, they can output the desired properties based on several structural and operating parameters with reasonable accuracy, which greatly reduces the time cost in engineering applications. For example, the Leverett J correlation has been widely employed in petroleum engineering to correlate the water saturation and capillary pressure in porous oil sandstones using only porosity ϵ , permeability K , surface tension coefficient γ , and wettability angle θ .³⁰

3. Modern energy materials research: explosion of porous energy material types, powerful imaging and modelling techniques

Modern energy materials research was mainly marked by three key milestones. The first one was the boom in types of artificial

porous energy materials that were fabricated for various energy devices.^{31–38} The second was the first appearance of commercial scanning electron microscopy (SEM) that was proposed by the Cambridge Scientific Instrument Company in the early 1960s. The third was the first computational simulation that was implemented by a team at the Los Alamos National Lab in 1957 which opened opportunities to numerically model physical processes in porous energy materials.²⁶

In more recent years, the increasing awareness of the adverse environmental impact of the large-scale utilisation of fossil fuels has motivated the development of various green and advanced renewable energy technologies, including energy conversion, storage and transfer^{39–44} (Fig. 2). Among these various energy devices, porous energy materials play significant roles in providing storage sites for gaseous energy carriers, high-conductivity transport paths for charge, photons, heat and mass, catalytic reaction sites and mechanical support. For example, a variety of active metal ions/clusters and organic linkers can lead to the high porosity found in metal–organic frameworks

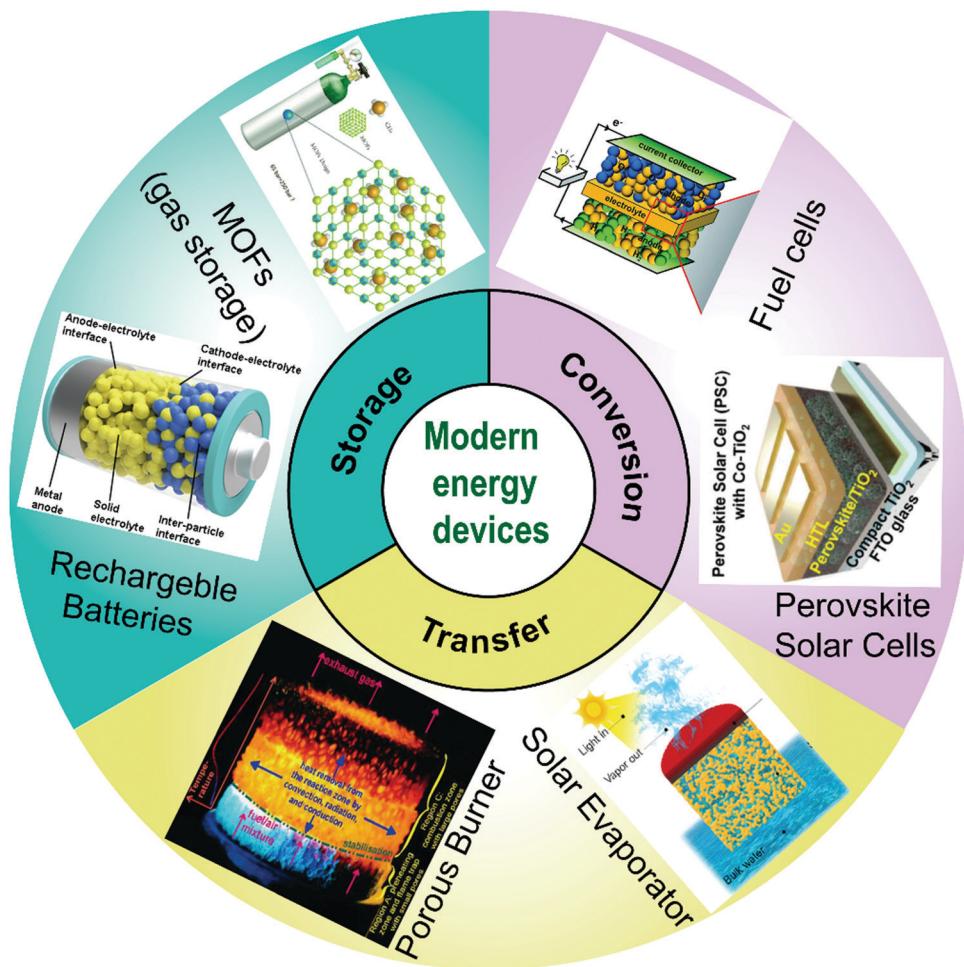


Fig. 2 Schematics of advanced energy devices that host modern energy materials.^{39–44} (The image of porous burner is reproduced from ref. 39 with permission from Elsevier. The image of rechargeable batteries is reproduced from ref. 40 with permission from Elsevier. The image of perovskite solar cells is reproduced from ref. 41 with permission from Wiley. The image of fuel cells is reproduced from ref. 42 with permission from Royal Society of Chemistry. The image of solar evaporator is reproduced from ref. 43 with permission from Elsevier. The image of MOFs is reproduced from ref. 44 with permission from Royal Society of Chemistry.)

(MOFs) and provide substrates for high-density gas storage.⁴⁵ In solar cell applications, the microstructures of the electron transport layer significantly affect the degree of perovskite infiltration, light trapping and harvesting, and charge injection/transportation.⁴⁶ Porous and hydrophilic exfoliated graphite and carbon foams are often combined in solar evaporators to transport water and insulate for heat loss.⁴⁷ To date, material digitalisation has primarily supported experimental design and optimisation and takes two forms: (i) digitalising the outputs of experimental characterisation techniques that often produce digital images of sample microstructures and (ii) simulation of the multi-physics processes to predict performance of complex microstructures that are often reconstructed by imaging techniques. The following sections will introduce the recent progress made in various imaging techniques for digitalising material microstructures (Section 3.1), in mathematical models for generating synthetic microstructures (Section 3.2) and in multi-scale modelling approaches for predicting material properties (Section 3.3).

3.1 Advanced imaging techniques

There are several typical imaging techniques that have been widely used to reconstruct digital 2D and 3D microstructures of porous energy materials at high resolution, *i.e.* SEM^{48–51} and atomic force microscopy (AFM) for 2D imaging,^{52–54} while focused ion beam-scanning electron microscopy (FIB-SEM),^{55–60} X-ray computerised tomography (CT),^{61–66} and magnetic resonance imaging (MRI) microscopy are used for 2D and 3D imaging.^{67–69} It is noted that neutron imaging has been popular in 2D imaging, however, it is mainly employed to probe fluid dynamics in fuel cells, material phases and magnetic structures or phase transitions.^{78–80} Fig. 3 highlights these techniques with their operating principles and imaging results. In some cases, temporal details are needed in addition to spatial information, to understand the evolution of chemical, physical and topological parameters. Thus, 4D (3D plus time or chemical species, or more generally $nD + 1$) reconstruction techniques are used, such as by individually employing X-ray CT

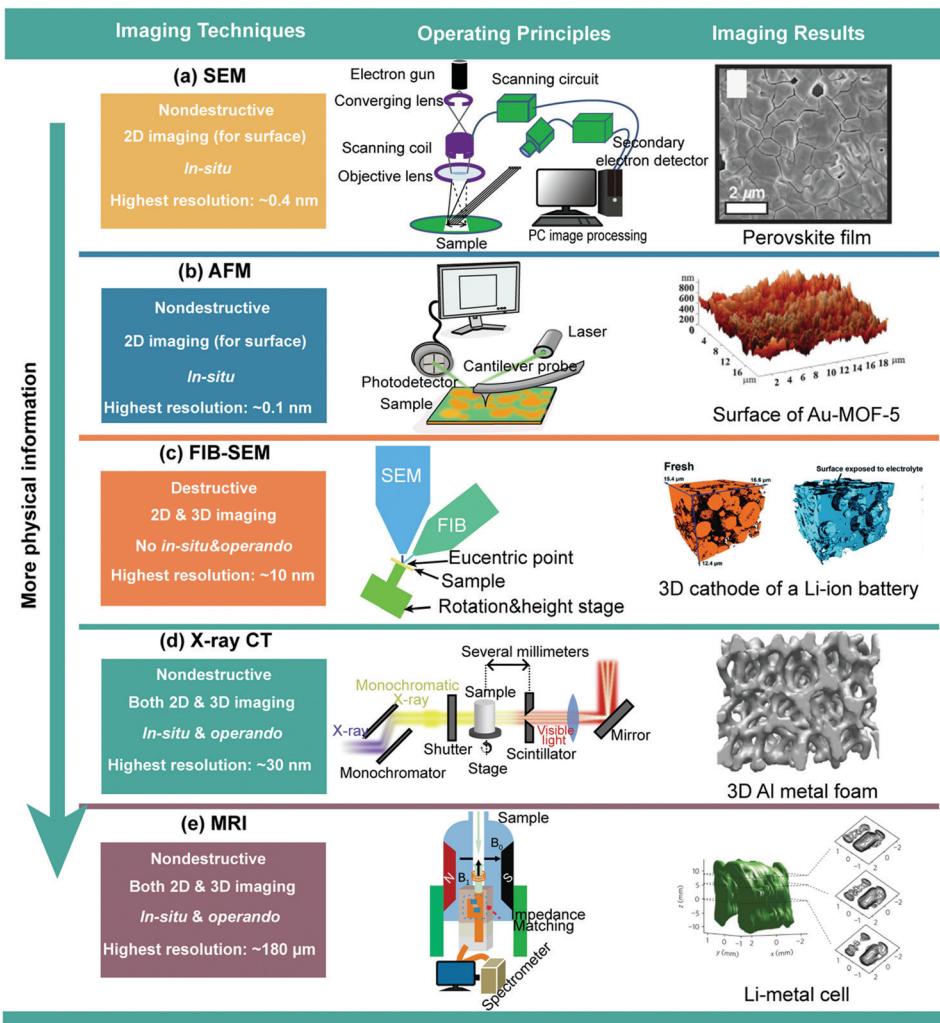


Fig. 3 Typical imaging techniques for microstructure reconstruction of porous energy materials. (a) Schematic of SEM and micrograph of a perovskite solar cell⁵¹ (the perovskite film is reproduced from ref. 51 with permission of the American Association for the Advancement of Science); (b) schematic of AFM and 2D images of Au-MOF-5⁵⁴ (the surface of Au-MOF-5 is reproduced from ref. 54 with permission of Elsevier); (c) schematic of FIB-SEM and 3D images of the porous cathode of a Li-ion battery⁵⁸ (the FIB-SEM images are reproduced from ref. 58 with permission from Royal Society of Chemistry); (d) schematic of X-ray CT⁹⁷ and images of 3D metal foam⁹⁹ (the images are reproduced from ref. 99 with permission of Elsevier). (e) Schematic of MRI and scanning images of a Li-metal cell⁶⁹ (the Li metal cell is reproduced from ref. 69 with permission from Springer Nature).

and MRI^{67–69,81–88} or by integrating imaging techniques with other techniques such as synchrotron X-ray diffraction (XRD).^{89,90}

The related work will be introduced in Section 3.1.3.

3.1.1 2D imaging

(a) Scanning electron microscopy (SEM). SEM scans the surface of porous energy materials with a focused beam of electrons (Fig. 3a). SEM also allows *in situ* observation of material surfaces. The highest resolution SEM can reach is ~0.4 nm⁹¹ and a general SEM image usually takes several seconds to complete. SEM can, however, only observe a topographical surface and does not reveal information about crystal structures, which is the scope of transmission electron microscopy (TEM)⁹² and electron back-scattered diffraction (EBSD).⁹³ In contrast to SEM, the electrons in a TEM system pass through the sample (albeit a very thin slice) to determine the inner structures of the sample. EBSD usually is conducted by using a SEM equipped with an EBSD

detector. Fig. 3a shows an SEM image of a perovskite film in a perovskite solar cell.⁵¹

(b) Atomic force microscopy (AFM). AFM also targets the surface of porous energy materials and possesses high resolution of up to ~0.1 nm. Fig. 3b shows an AFM image of Au-MOF-5.⁵⁴ However, AFM only provides limited information due to its limited vertical range, magnification range and potential damage to the sample. To extract more information from materials, it is common to integrate AFM with other imaging techniques such as SEM.⁹⁴

Once images of the surface are captured by one of these techniques, digital image processing methods must be deployed in order to extract microstructural information from the images. For example, successful image segmentation, the identification of particular structures or objects in the image, is

important to guarantee data accuracy; the segmentation algorithm and its input parameters must be carefully chosen to reduce information loss. Several powerful software tools such as ImageJ and Fiji provide various image segmentation methods including thresholding, watershed and so on.

3.1.2 3D imaging

(a) *Focussed ion beam – scanning electron microscopy (FIB-SEM)*. FIB-SEM is a destructive technique which combines FIB and SEM techniques together. A focused ion beam uses a magnetic lens to focus a beam of ions such as Gallium ion (Ga^+), onto a very small area on the surface of the material. This results in the stripping, deposition, implantation, cutting and modification of the material. In recent years, FIB has been used to process materials, combined with SEM and other high-power electron microscopy methods, to analyse the 2D and 3D microstructures of meso-scale and nano-scale materials. In the operation of a FIB-SEM, the sample is cut into several hundreds of layers by the milling of FIB and each layer is then imaged by SEM to a very high resolution. By stacking these 2D images, the final sample microstructures are reconstructed. Fig. 3c shows a schematic of how the FIB-SEM reconstructs the porous cathode of a Li-ion battery towards the final 3D image data.⁵⁸ FIB-SEM is ideally suited for imaging micron and submicron scale to a resolution of about 10–15 nm. The limitations of this technique make it difficult to image the structure that is smaller than 5–10 nm.⁵⁹ Users should be cautious when determining transport properties based on the reconstructed microstructures from FIB-SEM. This is because the field of view in FIB-SEM is limited to around $40 \times 40 \mu\text{m}$, leading to the reconstructed domain being possibly smaller than the representative volume elementary (RVE) of the porous material. For example, Kelly *et al.*⁶⁰ concluded that the permeability and pore connectivity calculated from a 3D shale domain below $\sim 5000 \mu\text{m}^3$ are not accurate due to the computational domain being too small. They suggested a broad ion beam SEM (BIB-SEM) which can provide a larger field of view than FIB-SEM. Generally, FIB-SEM requires large amounts of time for milling that limits its imaging speed. Another limitation of FIB-SEM is its relatively poor resolution along the milling direction compared to the resolution in the other two directions. In recent years, helium ion beam (HIB) has been employed to accelerate milling and improve spatial resolution.⁹⁵ It is worth noting that both FIB-SEM and HIB-SEM will affect the original material structures when cutting slices.

(b) *X-Ray computerised tomography (X-ray CT)*. As opposed to FIB-SEM, X-ray CT is a non-destructive imaging technique, allowing quantitative or qualitative insight into complex microstructures of various porous energy materials across multiple length scales. There are various classifications for X-ray CT such as attenuation contrast tomography and phase contrast tomography. Because of the undesired edge enhancement effects in attenuation contrast tomography, X-ray phase contrast tomography is most frequently used. When the X-ray beam penetrates the samples at different angles, a series of 2D projections are generated and converted into 2D grey-scale reconstruction images based on

the Fourier slice theorem and a user-defined or open source library such as ASTRA.⁹⁶ Finally, these 2D images are stacked in series and 3D reconstructed microstructures are formed to describe the different phases after image segmentation. A typical X-ray CT experimental system is shown in Fig. 3d.⁹⁷ Generally, X-ray CT involves a wide range of spatial resolution from a few nanometres to tens of micrometres *via* synchrotron X-ray CT,⁶¹ X-ray micro CT,⁹⁸ X-ray nano CT⁶² and so on. One of the biggest advantages of X-ray CT compared with FIB-SEM is that X-ray CT is suitable for reconstruction of a large RVE with a low time cost. A typical example is the metal foam that has been extensively reconstructed *via* X-ray CT technique (Fig. 3d).⁹⁹ Therefore, in summary, the trade-off between resolution and imaging volume should be carefully considered when obtaining the target morphological characteristics.

(c) *Magnetic resonance imaging (MRI)*. Although MRI is typically associated with medical research and diagnosis, it is a promising tool for non-invasively probing the complex chemical and physical, spatially and temporally varying structures in energy devices. MRI uses the magnetic field gradients to spatially locate nuclei by making their precessional frequency dependent on position (Fig. 3e).¹⁰⁰ However, performing MRI experiments on the electrically conductive microstructures that typically exist in energy devices can lead to radio frequency losses and heating of components. To overcome this issue, several improvements have been proposed to achieve the measurement of metal microstructures. For instance, Chang *et al.*⁶⁷ and Bhattacharyya *et al.*¹⁰¹ successfully achieved imaging of lithium dendrites by measuring changes in the intensities and frequencies of Li-metal signals due to skin-depth effects and susceptibility shifts.^{67,101} Other indirect methods have also been proposed. Ilott *et al.* imaged the induced magnetic field produced by the cell itself and then related it to processes occurring inside the cell.⁶⁸ Fig. 3e shows a 3D MRI image of a Li-metal cell.⁶⁹ As is the case with X-ray CT, MRI is non-destructive, although it has a lower resolution than X-ray CT because of the inherent insensitivity of the lithium isotopes and relaxation phenomena. The resolution for a typical MRI scanner for solid-state materials is $\sim 180 \mu\text{m}$.^{67–69} Recently, several groups have achieved high-resolution nano-scale NMR (spatial resolution $< 10 \text{ nm}$).^{70–72} However, these techniques require special facilities, such as the use of nitrogen-vacancy centre on the tip of a scanning probe microscope.⁷⁸ MRI has also been employed to determine ionic transport in supercapacitors but no electrode microstructures have been imaged.^{73,74}

Notably, it is difficult to extract all the desired information from porous energy materials by using one single technique. In reality, multiple techniques are often tightly coupled to obtain more spatial and chemical information. For instance, Daemi *et al.* employed X-ray CT, XRD and FIB-SEM to analyse the morphology changes of microstructures and crystallographic changes in uncycled and cycled $\text{LiNi}_{0.33}\text{Mn}_{0.33}\text{Co}_{0.33}\text{O}_2$ electrodes.¹⁰² MRI is usually used in conjunction with nuclear magnetic resonance (NMR) spectroscopy to provide both spatial and chemical information.⁶⁹ In order to characterise porous



energy materials with multiple functions, other testing methods are usually coupled with imaging techniques to reveal pore-scale material behaviours. For instance, mesoporous TiO_2 can be found in a wide range of energy applications such as dye-sensitized and perovskite solar cells, light emitting diodes, photocatalytic hydrogen generation and so on.¹⁰³ Nong *et al.*¹⁰⁴ tied high-resolution TEM, XRD, X-ray photoelectron spectroscopy and electron paramagnetic resonance spectroscopy together to evaluate the microstructures of mesoporous Ru-doped TiO_2 and its activity for electrocatalytic hydrogen evolution reaction.

3.1.3 4D imaging. 4D imaging techniques provide additional temporal or chemical information over and above the 3D spatial information. To probe the realistic evolution of microstructures or chemical species, *operando* devices must be prepared. Although definitions of the terms “*in situ*” and “*operando*” vary, here we use “*in situ*” to refer to measurements in fully assembled energy devices that can either operate or remain static. “*Operando*” refers to measurements made in service. For imaging in 3D plus chemical species, both *in situ* and *operando* devices are appropriate, while for 3D plus temporal evolution, only *operando* devices are suitable. Owing to the unique requirements of *in situ* and *operando* measurements, FIB-SEM was excluded from 4D imaging due to its intrinsic destructiveness, while the need for a vacuum and the sample damage prevent the utilisation of SEM and AFM. Environmental SEM has, however, been used in a 2D + 1 mode for the investigation of chemical distribution on surfaces,^{75–77} such as crack and liquid water distribution⁷⁶ in electrodes.⁷⁷ Non-destructive X-ray CT and MRI have therefore been the main techniques for 4D imaging.^{81–88} Wu *et al.* employed 4D soft X-ray

CT at multiple X-ray energies to resolve spatial ionomer distribution in the 3D reconstructed catalyst layer in a proton exchange membrane fuel cell (PEMFC),⁸⁴ see Fig. 4a. In the work of Wu *et al.*, the 4th dimension is the spatial ionomer distribution among catalyst particles, which was obtained by image contrast based on the different X-ray absorption of materials. In order to reduce the damage from the X-rays on the ionomer distribution, multi-set tomograms are used to guide dose reduction to achieve damage reduction (<5%). Besides characterizing microstructures, 4D X-ray CT has also been used to visualise the dynamics of multiphase flows in microstructures, such as water clusters in the cathodes of PEMFCs.⁸³

Another recent popular 4D imaging technique is *operando* MRI due to its quick scanning, high sensitivity and reasonable spatial resolution. Ilott *et al.* employed both MRI and NMR spectroscopy to resolve temporal and spatial evolution of Li dendrite microstructures in Li–metal batteries,¹⁰⁵ as shown in Fig. 4b. Their techniques offer 180 μm isotropic spatial resolution within temporal resolution of 16 min 40 s. When employing 4D imaging techniques to probe chemical concentration dynamics, the sensitivity to chemical concentration is critical. For example, a sufficient sensitivity (Li⁺ concentration resolution <10 mM) and fine temporal (faster than 1 s per frame) and spatial resolution (finer than 1 μm) are required to characterise 3D ion transport in electrolytes.¹⁰⁶ Fig. 4c shows Li⁺ concentration distributions in electrodes of a Li-ion battery that was probed by Raman scattering microscopy.¹⁰⁶

In summary, 4D imaging data has the potential to offer new insight into the phase transformation, particle/electrode morphology effects, nucleation propagation and particle–particle

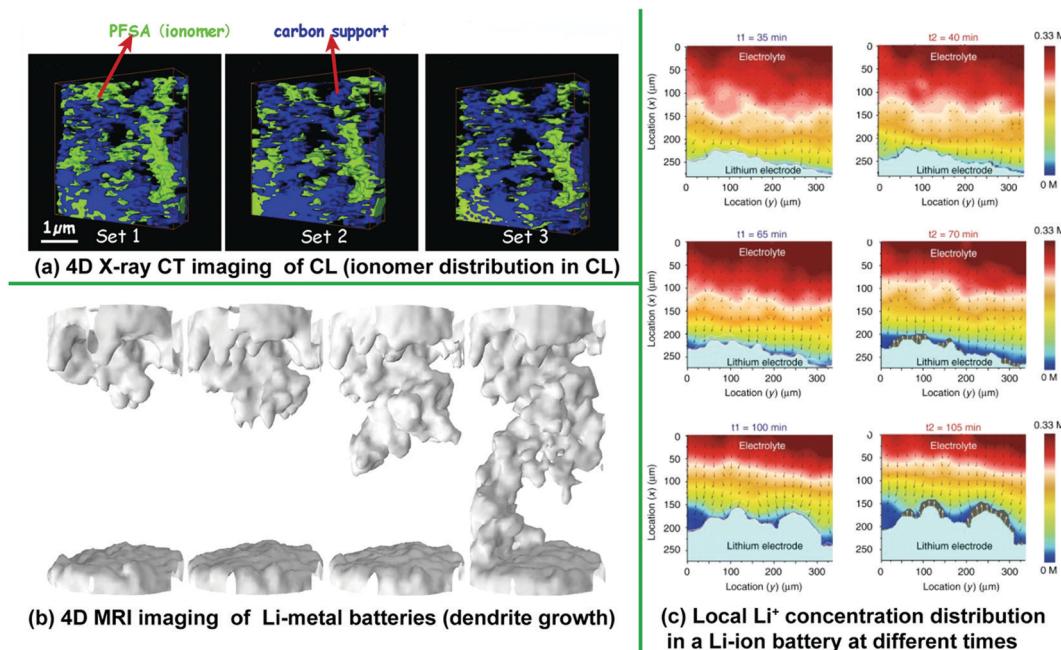


Fig. 4 (a) 4D imaging results of ionomer distribution in the CL of a PEMFC⁸⁴ (the image is reproduced with permission from Elsevier), (b) Li dendrite growth in Li–metal batteries¹⁰⁵ (the image is reproduced with permission from National Academy of Sciences) and (c) local Li⁺ concentration distribution in a Li-ion battery at different times¹⁰⁶ (the image is reproduced with permission from Springer Nature).



interactions, which can then be applied to advance high-performance porous energy materials. In order to track more multi-scale information on porous energy materials, 5D (three spatial and one temporal dimension, plus chemical composition) imaging has been further proposed to enhance the understanding of chemical evolution leading to particle formation within the internal porous structures and the impact of those particles on performance.¹⁰⁷

3.2 Mathematical models for generating microstructures of porous energy materials

The diversity of the microstructures reconstructed by the digital imaging techniques presented in Section 3.1 depends on the numbers of specimens provided, leading to a huge imaging cost for discovery and optimisation of microstructures. Besides, the reconstructed microstructures reflect, in turn, few formation processes that led to the specimens. In this regard, mathematical models specialised for generating microstructures are advantageous because their governing algorithms not only mimic the fabrication processes and synthesise vastly different microstructures by adjusting several important physical parameters, but are also fast and low-cost. Based on the various algorithms employed to construct these synthetic microstructures, we categorise the approaches into the stochastic fibre stacking method, the stochastic grain packing method, the simulated annealing method (SAM) and the multipoint statistical (MPS) method.

3.2.1 Stochastic fibre stacking method. The stochastic fibre stacking method is popularly employed to generate the microstructures of gas diffusion layers (GDL) of fuel cells and electrolyzers.^{108–112} The workflow of a typical stochastic fibre stacking method is depicted in Fig. 5a. Schulz *et al.* first introduced the stochastic fibre stacking method to generate synthetic microstructures of non-woven GDLs, where the pore size distribution and two-phase characteristics of their synthetic GDL agreed well with experimental one.¹⁰⁸ However, in their work, the binder and hydrophobic agent polytetrafluoroethylene (PTFE) were ignored. To further improve the approach of Schulz *et al.*,¹⁰⁸ Hao *et al.*¹⁰⁹ and Niu *et al.*¹¹⁰ incorporated the effects of PTFE into the synthetic GDL, whilst Hinebaugh *et al.*¹¹¹ and Burganos *et al.*¹¹² added binder and PTFE agents to the synthetic GDL. In these models, PTFE was considered by adding a PTFE sheath around fibres. Their model^{108–112} accounted for many parameters such as the heterogeneity and pore size distribution in the through-plane direction. Based on these improvements, more realistic GDL microstructures were constructed to help model and understand their multi-phase flow behaviour and to predict their transport properties. However, there are some challenges in stochastic fibre stacking methods. The first is that most models assume the fibres have identical diameter. The second is that additional topological descriptions are necessary to generate curved fibres, which complicates the model and increases the time cost. The final challenge is the low efficiency when these models are used to generate compressed fibrous samples because they need to be used in conjunction with other mechanical modelling methods such as finite element methods.

3.2.2 Stochastic grain packing method. Stochastic grain packing methods were first introduced to describe the microstructures of rocks and sandstones, and then further developed for granular porous media, such as electrodes of PEMFC,¹¹³ SOFC^{114–117} and Li-ion batteries.^{118–123} In the general stochastic grain packing methods, spherical particles of identical diameter are randomly placed in the given domain, and particles are allowed to overlap by a finite distance to ensure the connectivity of the sample. Once the desired porosity is reached, the algorithm ends. Additional procedures are sometimes necessary to eliminate unrealistic microstructures such as isolated particles. Fig. 5b shows a stochastic grain packing method for the anode of a Li-ion battery.¹²⁰ The random tessellation determines the distribution of grain shapes, sizes and locations. It is noted that general stochastic grain stacking models do not really capture the physics of the manufacturing process and few consider physical behaviour such as solvent mixing, sedimentation, drying and migration of particles induced by binder. Hannach *et al.*¹¹³ proposed a novel stochastic grain packing method to generate realistic micro-porous layer (MPL) structures in PEMFCs based on three physical input parameters: porosity, the diameter of carbon particles and the PTFE loadings which is generally used for hydrophobicity enhancement. Their synthetic MPL shows reasonable agreement with the experimental pore size distribution and transport properties (less than 10% error for effective diffusivity). One limitation of their model is that cracks formed by compression and cycling are not considered. This work was the first to consider both carbon particles and PTFE coatings in the synthetic MPL. This method has been also extensively employed to generate electrodes of SOFC,^{114–117} Li-ion batteries^{118–123} with various stochastic algorithms such as graph-based and Gaussian random field techniques. Based on these random graphs, the spatial distribution of different material grains is constructed.^{115,116} In the Gaussian algorithm, Gaussian random fields are introduced on the surface of particles to generate realistic particle geometries.¹²⁰ The electrode microstructures generated *via* these two methods not only visually fit the morphology characteristics of real electrodes, but also agree reasonably well in terms of porosity, effective electric and ionic conductivity. The main limitation of these stacking grain methods includes artificially determined overlap among particles, requiring additional experimental images for extracting morphological information and ignoring the heterogeneity of additives along the thickness direction on the surface of grains.

3.2.3 Simulated annealing method (SAM). The SAM is an optimisation method that mimics the slow cooling of metals. This process is characterised by a progressive reduction in the atomic movements that reduce the density of lattice defects until the lowest-energy state is reached. Inspired by this physical process, microstructure generation is treated as an optimisation problem where the objective is to achieve targeted microstructural properties that may be defined by some specified statistical characteristics of the target system. Grains in a computational domain randomly move to find a structure that satisfies specified statistical functions representing the target porous materials, see Fig. 5c, with gradually reduced scope for those movements as the



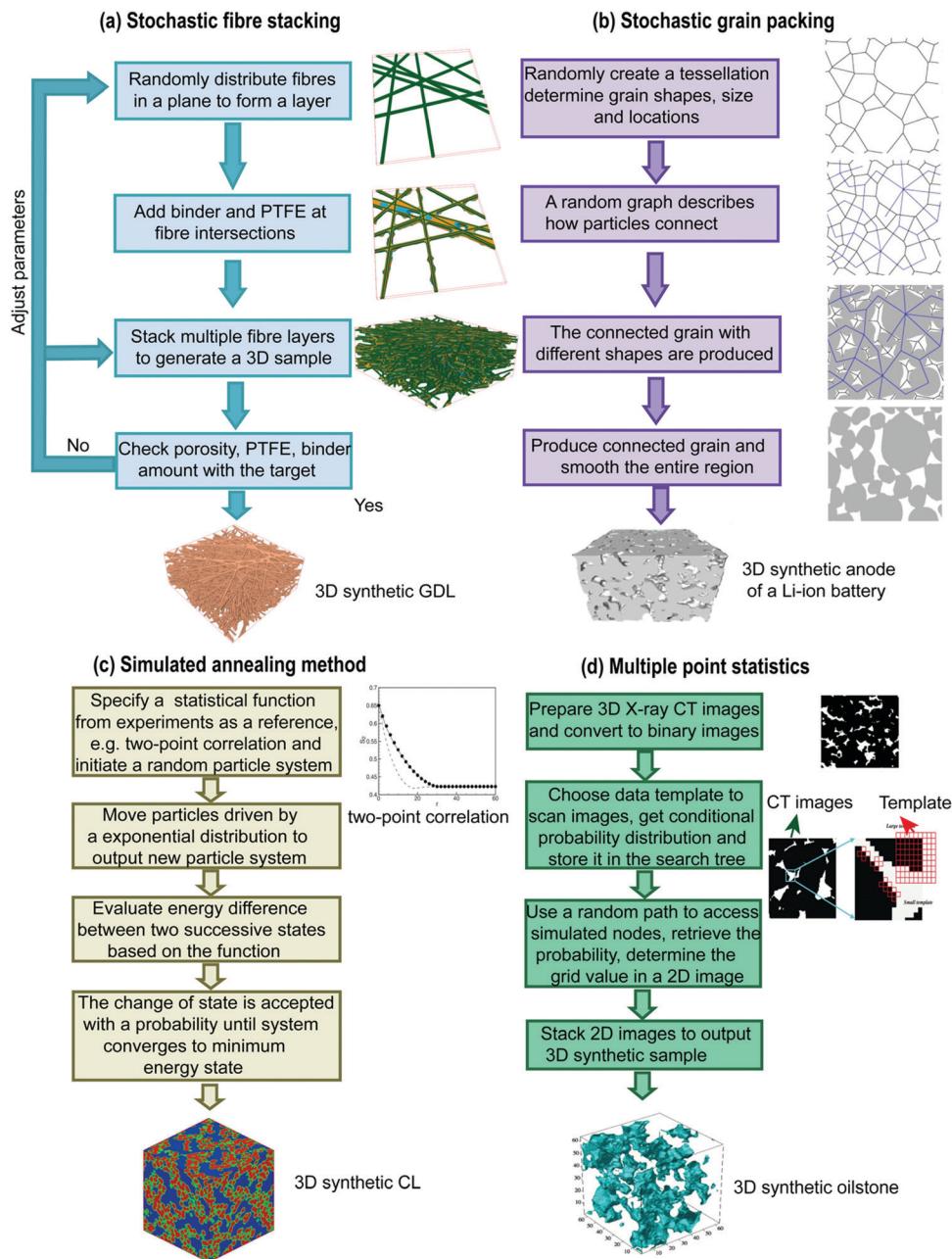


Fig. 5 Workflows of four stochastic microstructure generation algorithms. (a) Stochastic fibre stacking method¹³⁶ (the fibre images are reproduced with permission from Elsevier); (b) stochastic grain packing method¹²⁰ (the fibre images are reproduced with permission from Elsevier); (c) simulated annealing method¹²⁴ (the 3D synthetic CL is reproduced with permission from IOP publishing); (d) multiple point statistics method¹³² (the 3D synthetic oilstone is reproduced with permission from American Physical Society).

objective is approached. Kim and Pitsch *et al.*¹²⁴ pioneered the application of a SAM for the generation of CLs in PEMFCs and validated the predicted pore size distribution with experimental data. One of their synthetic CLs is shown in Fig. 5c. Consequently, Jiang *et al.* introduced the SAM to the electrode generation of PEMFCs¹²⁵ and Li-ion batteries.^{126–130} Yin *et al.*¹³¹ employed a SAM to generate Gosford sandstone and compared the synthetic sample with experimental images from micro-CT. In these works, they showed that the SAM could simulate and replicate statistical spatial information in the synthetic microstructures.

The significant limitation of SAM is that its computational cost is higher than other methods due to the iterative computation to search for the minimum energy state.

3.2.4 Multiple-point statistics (MPS). MPS is a high-order statistics method, which describes the statistical relation between multiple spatial locations based on training 2D images from SEM or micro X-ray CT techniques. This method is more advantageous in reproducing the long-range connectivity of the pore space than stacking grain packing methods because MPS is based on a set of training images and employs multi-point

Table 1 Pros and cons of various microstructure generation methods

Methods	Pros	Cons
Stochastic fibre	<ul style="list-style-type: none"> ✓ Pixel-based or geometry-based ✓ Easy to program ✓ Easy to generate vastly different samples 	<ul style="list-style-type: none"> • Fibres are same diameter • Complicated topological descriptions are necessary for curved fibres • Long time for generating compressed samples (need conjunction with other mechanical modelling such as finite element method)
Stochastic grain	<ul style="list-style-type: none"> ✓ Only requires simple stochastic algorithms ✓ Synthetic microstructures display inhomogeneous properties similar to realistic ✓ Pixel-based or geometry-based ✓ Easy to program ✓ Easy to generate vastly different samples ✓ Allows diverse grain shapes to mimic realistic particle geometries ✓ Synthetic samples display visually and quantitatively appropriate characteristics ✓ Pixel-based 	<ul style="list-style-type: none"> • The overlap between grains is manually determined • Some methods require experimental images to extract morphological information • The heterogeneity of binder, ionomer and agents in thickness was ignored
Simulated annealing	<ul style="list-style-type: none"> ✓ Physically meaningful governing algorithms ✓ Good pore connectivity ✓ Wide applications (including fibrous and granular materials) ✓ Pixel-based 	<ul style="list-style-type: none"> • Computational cost is higher than other methods (iterative computation is necessary)
Multi point statistical	<ul style="list-style-type: none"> ✓ Physically meaningful governing algorithms ✓ Displays good long-range connectivity 	<ul style="list-style-type: none"> • Requires experimental images to extract statistical information • Intensive computation • No consideration of the formation processes of microstructures

statistical functions to extract spatial connectivity and variability. To obtain a faithful local percolation probability, a huge set of training sample images is necessary. This method has been widely adopted to generate the microstructures of sandstones and rocks.^{132–135} The workflow of a typical MPS method is shown in Fig. 5d. Okabe and Blunt¹³² proposed an MPS method to generate 3D pore structures of the Fontainebleau, Berea sandstones and carbonate rock based on 3D micro-CT images. The statistical characteristics of Berea sandstone such as the autocorrelation function and fraction of percolating cells were validated. It is noted that a huge computational cost is required to extract local percolation probability in a 3D micro-CT image. To reduce the computational cost, Okabe and Blunt¹³³ further proposed an MPS method to generate 3D pore structures by using 2D thin slices as training images. Hajizadeh *et al.*¹³⁴ further reduced the computational cost by using a single normal equation simulation (SNESIM) algorithm for the successive 2D MPS simulation. These works were limited by the hardware of computers at the time. Recently, huge progress in GPUs and computer memory allows high-efficiency 3D MPS simulations on 3D micro-images again. Wu *et al.*¹³⁵ adopted both 3D MPS and SNESIM algorithm to generate two kinds of oil sandstones. Therefore, MPS methods could be employed for any porous media if 2D or 3D micro X-ray CT or SEM images are provided. It is noted that MPS requires experimental images for extracting statistical information which greatly increases the computation time. Besides, MPS mainly mimics the realistic microstructures in final form, but does not consider the

formation processes of microstructures. Table 1 summarises the pros and cons of four kinds of methods.

3.3 Structure–property–performance modelling approaches

The performance and functionality of porous energy materials is governed by physical processes occurring on a range of length scales from the atomic to the macroscopic and similarly for temporal scales. However, a great deal of insight has been gained by modelling such processes separately at different temporal and spatial scales, and in some cases combining models of a number of physical processes (multi-physics) at a particular scale.^{137–141} In this section, these physical modelling techniques are presented; a summary of them is shown in Fig. 6.

3.3.1 Atomic scale models. Atomic scale models are specialised to model physical behaviours of porous energy materials at the scale of atoms (10^{-10} m) and even subatomic scale. Temporal scales are still very short (\sim several ns), constituting one of the limitations of these methods. Two types of models have been developed for this purpose, *i.e.* density functional theory (DFT) and molecular dynamics (MD).

DFT is specialised to model subatomic scales (electrons) and involves quantum-mechanical theories for various types of molecular interaction to reveal complex physical behaviours at an atomic scale. DFT requires several parameters as inputs such as the coordinates and identities of the atoms in the material within a repeating lattice, the exchange–correlation functional, parameters and algorithms for numerical and iterative convergence



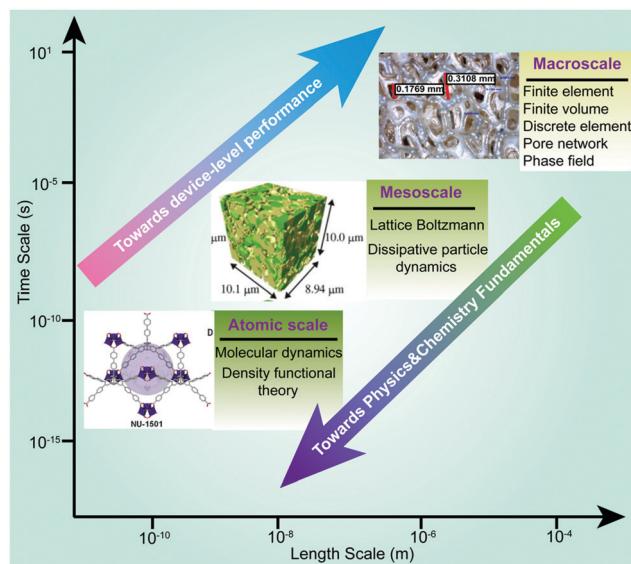


Fig. 6 Various physical models for capturing multi-physics dynamics in porous energy materials (images are reproduced from ref. 5, 57 and 138 with permission from the American Association for the Advancement of Science and Elsevier).

and so on.¹⁴² The basic properties output by DFT include electronic charge density, total energy, magnetic configuration and electronic band structure. Post-processing of output data from DFT calculation is important to derive useful properties of porous energy materials. A variety of material features such as elasto-mechanics, electronic structure, charge density and electrostatic potential, opto-electronic properties, wave function and catalytic activity are calculated based on various theoretical formulae. Recently, DFT has been applied to model catalysts, perovskite solar cells and MOFs. Wang *et al.*¹⁴³ employed DFT to study the effluence of Eu³⁺–Eu²⁺ ion pairs on the reaction energies of the redox reaction between Pb⁰ and I⁰, lattice stability, and energy band structure in perovskite solar cells, as shown in Fig. 7a. Masoud *et al.*¹⁴⁴ conducted a DFT study to understand the optoelectronic properties of S-doped MoO₃ and O-doped MoS₂ bulk systems, with insights into the effective mass of electrons and holes, electron and hole mobilities, and exciton binding energy. The excited-state opto-electronic properties can be obtained by using time-dependent DFT, with Solomon *et al.*¹⁴⁵ using this to predict nonlinear optical and optoelectronic properties of vinyl coupled triazene chromophores such as absorption maxima, electronic transition energies and oscillator strength. Whang *et al.*¹⁴⁶ employed a time-dependent DFT (TD-DFT) to calculate singlet and triplet excitations using the B3LYP level of theory and they successfully predicted the absorption spectra of ReTPS and Lehn catalysts which are used for enhancing the reduction of CO₂ to CO. Chen *et al.*⁵ performed DFT calculations to extract the pore-size distribution of ultra-porous MOFs namely Nu-1501-Al, and investigated their structural characteristics and their effect on the storage capacities of methane and hydrogen. DFT has also been extensively employed in other porous energy material calculations such

as Li-ion electrodes¹⁴⁷ and catalysts of fuel cells.¹⁴⁸ An in-depth review of DFT calculations of porous energy materials has been found in Jain *et al.*¹⁴⁹ Generally, DFT calculations cover the whole periodic table, provided the adequate pseudo-potentials have been developed. However, the choice of exchange–correlation functional largely determines the accuracy of the DFT calculation. Other limitations of the standard DFT are the difficulty in modelling weak interactions, long-time dynamics and properties of excited states.¹⁴⁹

Compared with DFT, MD provides more direct insight into complex mechanisms at the molecular scale, allowing users to model interactions among thousands of molecules. This method relies on the solution of the Newton's equations of motion for all atoms in order to extract path-dependent processes of materials. Fan *et al.*^{150,151} employed MD to predict the electrochemical surface area of porous CLs in PEMFCs under different humidity conditions, and the oxygen transport/thermal conductivity for a perfluorosulfonic acid membrane, as shown in Fig. 7b. MD has also been employed to investigate Li⁺ transport within a solid electrolyte of Li-ion batteries.^{152,153} It is noted that high-performance computational facilities are often required for simulations of sufficient size ($\sim 10^7$ + molecules). Besides, MD models can only simulate physical processes at nanosecond time scales, which requires special treatment of the initial fields to accelerate the simulation. For example, the air pressure needs to be increased to tens of MPa to increase the number of oxygen molecules in a small computational domain.^{154,155} Thus, it has been challenging to validate MD model results with experimental data due to the small computational domain and short physical time. MD simulations show good precision when predicting material properties (e.g. density) and transport properties (e.g. diffusion coefficient), with errors less than 5%.¹⁵¹ The material properties predicted by MD models greatly contribute to the in-service model of energy devices by indicating relationships between different physical variables, for example, the quantitative relationship between relative humidity and electrochemical surface area in fuel cells.¹⁵⁰ As classic MD is designed for simulations at the time scale around tens of ns, for slower processes outside this timescale range, kinetic Monte Carlo (KMC) models have often been employed as well.¹⁵⁶

3.3.2 Mesoscopic models. Mesoscopic models are designed to address problems at mesoscopic length scales (100 nm–100 μm) but where continuum assumptions are not suitable. There are two commonly used mesoscopic models namely the lattice Boltzmann model (LBM) and dissipative particle dynamics (DPD).

The LBM is a popular modelling approach originally for complex fluid systems in computational physics based on microscopic particle models and mesoscopic kinetic equations¹⁵⁷ and has now been extended to model various physical transport processes in porous energy materials.¹⁵ LBM is advantageous in dealing with complex flow and transport problems in complex porous domains due to its unique particulate nature and local dynamics. Numerous LBM studies have been conducted to predict various properties of porous energy materials. Chen and Tao^{158,159} have employed LBM to predict effective permeability

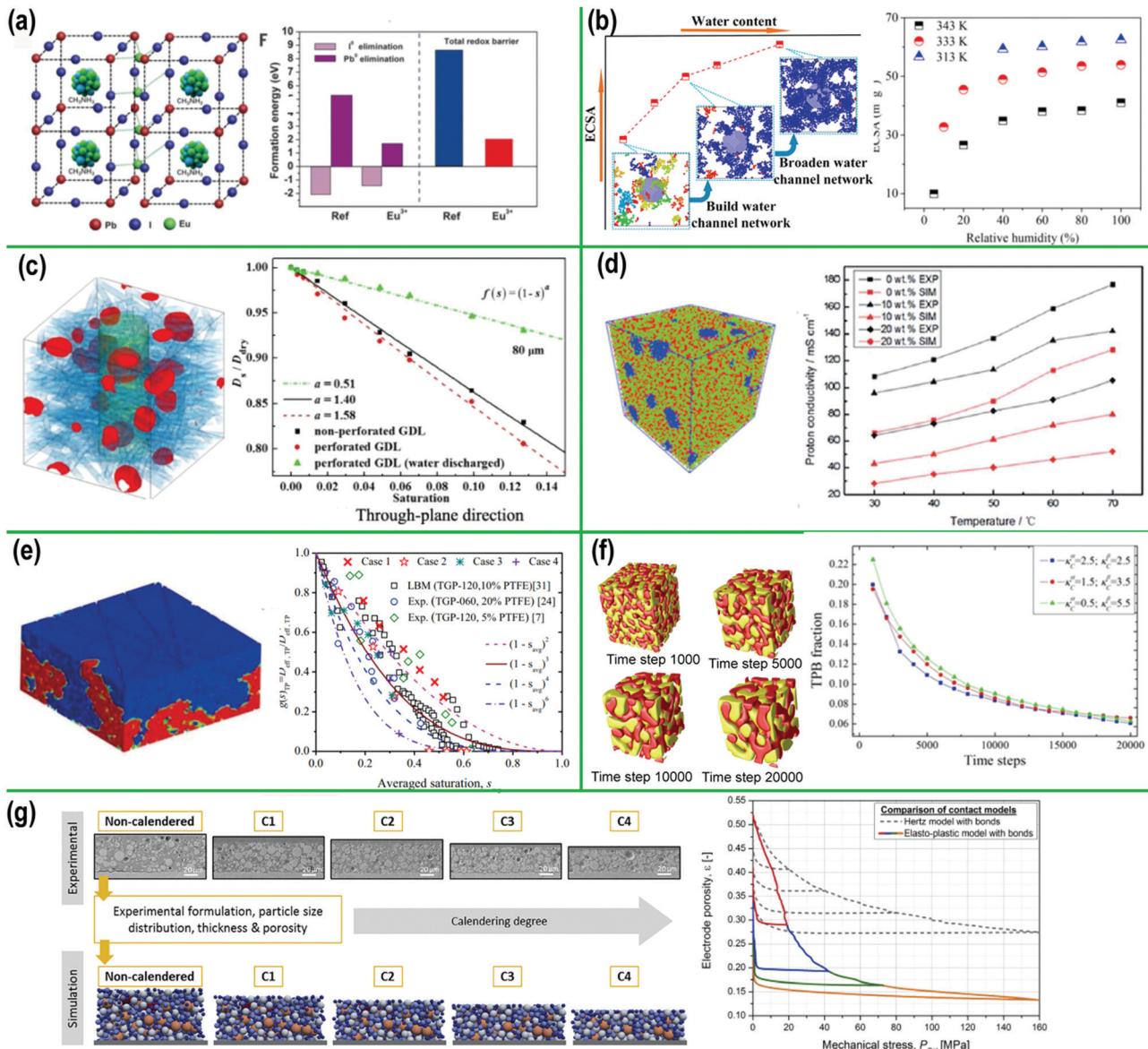


Fig. 7 Results of various physical models that were used to predict properties of porous energy materials. (a) Eu clustering-layer-incorporated and formation energies in perovskite solar cells (DFT)¹⁴³ (reproduced with permission from the American Association for the Advancement of Science); (b) structures of CL and electrochemical surface area (MD)¹⁵⁰ (reproduced with permission from ACS Publications); (c) effective diffusivity of perforated GDLs (LBM)¹⁵⁸ (reproduced with permission from Elsevier); (d) structure of blend membrane and proton conductivity (DPD)¹⁶³ (reproduced with permission from Royal Society of Chemistry); (e) effective diffusivity of wet GDLs¹⁶⁴ (reproduced with permission from IOP Publishing); (f) structural evolution of an SOFC electrode and three-phase boundary fraction (PFM)¹⁶⁶ (reproduced with permission from AIP publishing); (g) structural evolution of electrode of Li-ion battery in calendaring process¹⁶⁷ (reproduced with permission from Elsevier).

and diffusivity of porous shales and the GDLs of PEMFCs. The predicted diffusivity of a wet GDL is shown in Fig. 7c. In their models, imaging techniques and mathematical models offered computational microstructures for LBM calculations. Similar LBM works have also been conducted to predict effective thermal conductivity, electric and species transport properties of electrodes of Li-ion batteries.^{160,161}

DPD is an off-lattice mesoscopic technique that involves a set of particles moving in continuous space and discrete time. Particles represent whole molecules or fluid regions, rather than single atoms, and atomistic details are not considered

relevant to the processes addressed. The particles' internal degrees of freedom are integrated out and replaced by simplified pairwise dissipative and random forces, so as to conserve momentum locally and ensure correct hydrodynamic behaviour.¹⁶² The main advantage of this method is that it gives access to longer time scales and larger length scales than the most challenging conventional MD simulations. Simulations of polymeric fluids with length scales up to 100 nm for tens of microseconds are now common. Ma *et al.*¹⁶³ employed DPD to predict the transport properties and proton conductivities of the blend membrane of PEMFCs. Fig. 7d shows their



membrane structures and predicted proton conductivity from DPD modelling.

3.3.3 Macroscopic models. Macroscopic models are usually used to address engineering problems typically at large length ($>10^{-5}$ m) and time scales ($>10^{-6}$ s) where the continuum approximation holds. For these models, a representative region needs to be determined as the computational domain with a discretised mesh. Non-linear governing equations are then discretised in this domain and solved using initial and boundary conditions, once numerical schemes and solvers are specified. Diverse macroscopic models have been developed to tackle different problems such as fluid dynamics, mechanics, multi-phase flow and so on. For example, finite element models (FEM) and finite volume models (FVM) are specialised for structural analysis, heat transfer, fluid flow, mass transport, electromagnetics and so on. Fig. 7e shows predicted liquid water distribution and effective diffusivity of a wet GDL by using a FVM model.¹⁶⁴

There are several macroscopic models specialised for the structural evolution of porous energy materials to predict time-varying structural properties such as triple-phase boundaries (TPB), local porosity, surface area and so on. The first is the phase field model (PFM) which is a powerful mathematical model for solving interfacial problems. Due to their unique governing equations, derived from thermodynamic theory and named after Allen-Cahn, the PFM has significant advantages in predicting various behaviours of porous energy materials such as solidification, solid-state structural phase transformations, crack propagation, grain growth and so on.¹⁶⁵ Chen *et al.*¹⁶⁶ pioneered relevant work in the fundamental understanding of mesoscale microstructure evolution based on the PFM approach. They firstly employed the PFM to model the temporal evolution of microstructures during cell operation, and the TPB fraction (key reaction sites) of the electrodes in SOFCs. Fig. 7f shows the evolution of microstructures and TPBs of an SOFC electrode that were predicted by the PFM.¹⁶⁵ The second modelling method for structural evolution is the discrete element method (DEM) that is used to model the heterogeneity, discontinuity, motion, and large deformation of numerous particles or blocks. Therefore, DEM is popular not only in modelling particle fracture in electrodes of batteries, but also rock stability analysis. Fig. 7g shows a DEM model that was used to investigate the effects of calendering process on the microstructure and porosity of Li-ion battery electrodes.¹⁶⁷

4. The future of energy materials: digitalisation of porous energy material design and optimisation

The past and modern research reported in Section 3 has covered high resolution imaging enabling digital representation of structure, digitalised construction of *in silico* structures comparable to physical structures, and simulation of properties and physical phenomena in porous structures at a range of length scales and time scales. However, these have essentially supported the experimental design and optimisation of porous energy materials. Now,

digital techniques take the fore in design, optimisation and material discovery, through the application of additional digital methods, including artificial intelligence (AI), to address the challenges of current/previous approaches. However, there are four challenges still remaining. (1) High dimensionality. As mentioned above, energy materials often involve multi-physics phenomena which span multiple length scales. This high dimensionality requires optimisation approaches that have multiple criteria. (2) Multi-scale modelling. It has been challenging to fuse various computational approaches to modelling multi-physics behaviours into an integrated and high-efficiency computational platform due to the large number of numerical iterations required to connect the different temporal and spatial gaps. (3) Powerful graphics-processing ability. The microstructures of energy materials are generally stored in the form of digital graphics. Current graphics-processing algorithms are inadequate for microstructural design and optimisation because they show poor performance for material images construction and structure rendering. (4) Material discovery *via* high-throughput screening. In the digital materials space, tens of millions of candidate structures could be proposed so efficient screening and optimisation methods are even more necessary. At this stage, energy material digitalisation comes to the fore and is beginning to act as the primary tool for design and optimisation, taking the place of experimental trial and error. All material information, ranging from material structures, properties and even the performance in service, is shaped into virtual data spaces so that any information inquiry, data mining, feature extraction and relationship prediction can be efficiently made. Also, material digitalisation can accelerate novel design, microstructure creation and its optimisation using simulations of performance, all of which are *in silico* and with experimental validation and refinement.

With the rapid development of soft-computing techniques and advanced statistical theories, AI techniques are starting to play key roles in the digitalisation of porous energy materials. Recently, AI has been applied in various scientific and engineering areas including material science due to its abilities in data mining and feature extraction.^{168–175} In conjunction with the experimental techniques and mathematical models introduced in Section 3, AI techniques have been successfully employed in the study of energy materials. Various machine learning (ML) methods and advanced deep neural networks (DNN) have shown excellent performance in regard to material structure reconstruction and generation, and property and performance prediction, such as artificial neural networks (ANN),¹⁷⁴ support vector machines (SVM),¹⁷⁶ convolutional neural networks (CNN),^{177–182} generative adversarial neural networks (GANN)^{183–190} and so on. Fig. 8 shows a simplified schematic of how AI can work in combination with conventional techniques to promote the full digitalisation of porous energy materials. In the following sections, we present a thorough review of recent applications of AI techniques in microstructure reconstruction and generation, property prediction and performance modelling, especially addressing the applications of several deep learning methods such as GANN and CNN.



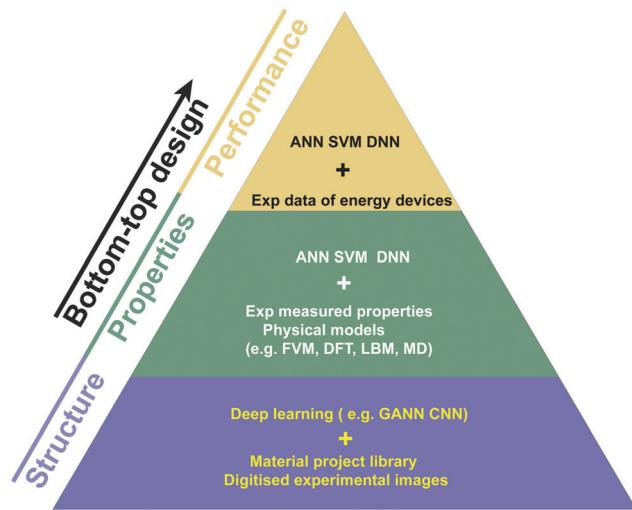


Fig. 8 AI-enabled structure reconstruction and generation, prediction of properties and in-service performance for porous energy materials.

4.1 Structure reconstruction and generation

As highlighted in Section 3, digital imaging processing is an important step to reconstruct or generate microstructures.^{108,113} The recently successful applications of AI in videos and graphics have opened new opportunities in microstructure reconstruction and stochastic generation based on learning from digitised experimental images of physical structures.^{177,178,183–190} In this respect, there are two popular AI techniques that substantially accelerate the development, *i.e.* convolutional neural networks (CNN) and generative adversarial neural networks (GANN). CNNs

can accelerate image segmentation and manipulation because of their powerful ability for feature extraction, leading to improved efficiency and accuracy of microstructure reconstruction. The main purpose of a GANN is to synthesise new microstructures based on characteristics learnt from training images. In some specialised GANNs, CNNs have been incorporated to improve their effectiveness.

(a) Convolutional neural networks (CNNs). As discussed in Section 3, before stacking numerous 2D images from FIB-SEM or X-ray CT into a 3D digital image, these grey-scale images usually need to be segmented into their respective phases. The accuracy of image segmentation directly impacts the pore-scale characterisation and related properties. However, all segmentation methods have user-selected parameters that result in biases, whereas a well-trained CNN can provide much more consistent image segmentation. CNNs use the mathematical principle of convolution to extract features from localised regions of the input data (often an image); this is achieved by applying a filter that takes data from a set of neighbouring datapoints, to each region of the data consecutively. By compiling a set of such filters and learning different parameters for these filters, the system builds a description of the data based on features associated with meaningful relationships between datapoints in the original data. CNNs are highly effective at extracting features from the data, particularly at different scales (based on the design of the filters). Niu *et al.*¹⁷⁸ employed a CNN called LeNet-5 to segment digital sandstone images from X-ray CT and SEM experiments and evaluated the results in terms of porosity, permeability and pore size distribution, resulting in greatly reduced variance in property prediction (reducing the variation of absolute permeability from

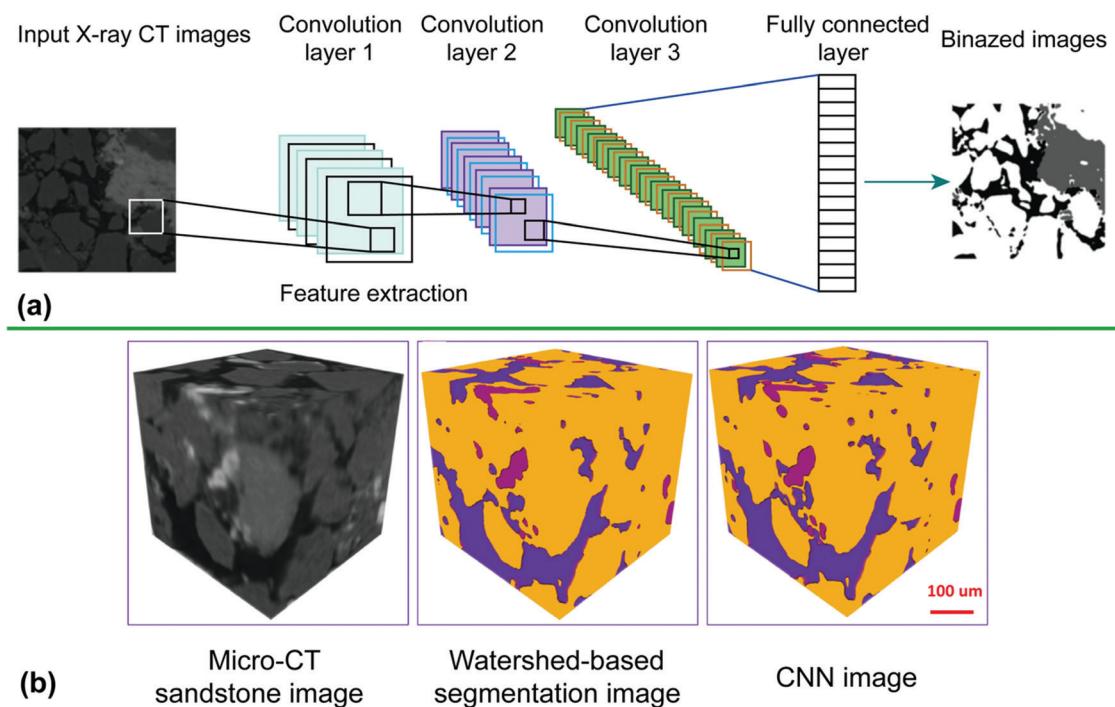


Fig. 9 (a) Schematic of the workflow of a typical CNN; (b) comparison of sandstone images from micro-CT, watershed-based segmentation and CNN¹⁷⁸ (reproduced with permission from Wiley).



64.9% to 14.2%) compared with watershed-based segmentation methods at different threshold settings. Fig. 9 shows the workflow of a typical CNN and the segmentation results of Niu *et al.*¹⁷⁸ Besides image segmentation, CNNs have also been employed to improve low-resolution images of datasets with a large field of view to produce high-resolution images with an equal size area of interest.¹⁸⁵

(b) Generative adversarial neural networks (GANNs). Conventionally, experimental imaging techniques can only provide digital samples with finite diversity and field of view, limited by the cost and device capability. On the other hand, mathematical models often simplify the structure and make assumptions for complex morphologies. Thus, developing microstructure generation methods that include merits of both experimental techniques and mathematical models is desired. Through this strategy, limited experimental images provide learning samples to the stochastic mathematical models, which can then generate lots of synthetic microstructures that have similar characteristics and properties to the real samples. To this end, GANNs have been employed in the area of microstructure generation due to their unique neural network architecture proposed by Goodfellow *et al.*¹⁸⁶ The structure of a typical GANN consists of two networks: a generator network and a discriminator network.

The generator is a function that is applied to a sample from a latent random space and creates a synthetic realisation. The discriminator's role is to determine whether a sample is part of the training image data or from the generator. The misclassification rate is then computed and is fed back to the generator for further improvement in the quality of the produced samples, towards fooling the discriminator. When a sufficient image quality is obtained, training will be stopped and the discriminator is discarded. Finally, the generator will be used to create new digital microstructures (Fig. 10a). The first example of microstructural reconstruction based on a GANN was introduced by Mosser *et al.*¹⁸⁷ They developed a GANN to reconstruct microstructures of digital sandstones based on X-ray CT images. With this trained GANN, realistic microstructures of sandstones could be generated efficiently. Fig. 10 also shows digital electrodes of Li-ion batteries and SOFC generated by this GANN. It is noted that both training data and output results were both 3D images, which require large memory and long computational time. In their subsequent work,^{188,189} they adopted a deep convolutional GANN (denoted DCGAN) to reconstruct microstructures of sandstones, multiphase electrodes of lithium-ion batteries and SOFC anodes. This DCGAN is advantageous because both generator and discriminator employ CNNs, which

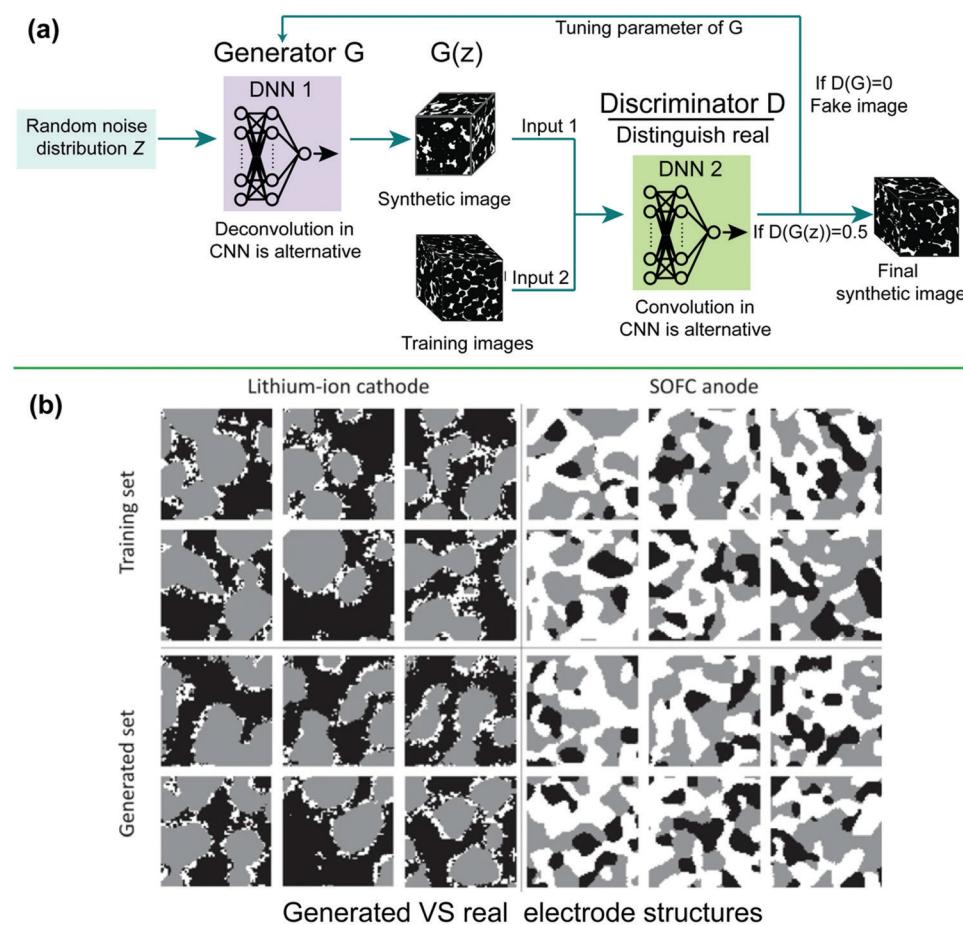


Fig. 10 (a) Schematic of the workflow for a typical GANN¹⁷⁸ and (b) comparison of generated and real multiphase structures for a Li-ion battery cathode and SOFC anode¹⁸⁹ (reproduced with permission from Springer Nature).

are powerful in extracting features from digital images compared with conventional GANs.

In the above GANN works, 3D images are usually required as training data. However, 3D image datasets are often unavailable for most researchers. In some complicated cases with extremely small scales, only 2D SEM images can be used to resolve the smallest features. Since GANs can be applied to both 2D and 3D images, a more convenient way to generate novel and realistic microstructures is to use only 2D images which are easy for users to assess. Valsecchi *et al.*¹⁹⁰ developed a GANN to generate 3D porous sandstones from 2D images. The key point of this approach is that the discriminator operates on 2D images, while the generator produces 3D images. Once a 3D image has been generated, a set of cross-sections are extracted and fed into the discriminator. Aside from 2D to 3D GANs, auto-encoders have also been introduced into the GANN to accelerate image data processing.¹⁸³ Apart from the utilisation of 2D images, the efficiency of GANs can be improved by optimizing their neural structures, since modern GANs are usually highly overparameterised. A recent pruning technique based on dynamic reallocation of non-zero parameters allows removal of a significant fraction of network parameters with little loss in accuracy.¹⁷⁹ He *et al.* employed similar techniques to prune filters, achieving a reduction of 52% in floating point operations.¹⁸⁰

4.2 Structure–property relationships

Understanding the structure–property relationships in porous energy materials is essential to accelerate design and discovery of new materials. However, properties of given microstructures are often predicted using empirical correlations or numerical modelling. Such methods are either specific or require high computational cost. The challenge is therefore to develop techniques to accurately and efficiently predict properties of new candidate structures using material databases that include both structure and property information (either experimental or modelled). In recent years, this challenge has prompted the exploration of data-driven models that can accurately and efficiently predict the structure–property linkage of porous energy materials. As one of the most exciting AI techniques, ML methods have been employed to predict properties of porous energy materials.^{171,191,192} Shandiz and Gauvin¹⁹³ employed various ML classification algorithms including linear, quadratic and shrinkage discriminant analysis, neural networks, support vector machines, k-nearest neighbours, random forests and extremely randomised trees to determine crystal systems of silicate-based cathodes in Li-ion batteries. In their work, the different compositions of the crystal system of silicate-based cathodes are taken from materials projects that offer an open web-based access to the calculated physical and chemical properties of known and predicted materials derived from DFT calculations of electronic structure.^{194–196} They found that random forests and extremely randomized trees gave the highest accuracy of prediction.

Since most microstructures of porous energy materials are based on digital images, the use of CNNs has gradually extended

beyond segmentation and reconstruction to the prediction of properties of porous energy materials like permeability of digital oilstones,^{181,197} structural properties (e.g. Young's modulus and Poisson's ratio),¹⁹⁸ effective thermal conductivity of porous carbon nanotubes¹⁹⁹ and effective diffusivity of general porous materials.²⁰⁰ As is well known, the key for DNNs in predicting the structure–property linkage is the availability of large training datasets. In the above works, digital images of microstructures can be from either experimental data like X-ray CT and SEM or mathematical reconstruction models. For the preparation of a properties database, numerical models such as LBM and FEM are often chosen because of their strong capability in modelling porous media.

Fig. 11a shows a schematic of a 3D-CNN used to predict structure–property linkages for porous materials.¹⁸² This 3D-CNN was employed to predict the effective elastic properties of high contrast composites based on their 3D microstructures as input. A high contrast composite implies that the elastic properties (e.g., Young's moduli) of the constituents are significantly different from each other. In this CNN, 3D images were firstly convolved with 32 filters in a convolutional layer with a filter size of $10 \times 10 \times 10$. The main purpose of the convolutional layer is to extract spatial features in the 3D images and learn parameters for a set of filters. It is noted that the size of the learned filters was informed by 2-point correlation statistics which define a characteristic length scale for the dominant features of the microstructure. Then the outputs of the convolutional layer were activated with the rectified linear unit activation function (ReLU) which is commonly used in CNNs. These outputs are then pooled into 8 values in an average pooling layer. The pooling layer was mainly used to reduce the number of parameters and computational time by down-sampling the outputs from the convolutional layers. Since there were 32 filters in this 3D-CNN, a total of 256 features were generated at the end of the pooling layer. A fully connected layer is usually at the end of the CNNs to derive the final classification or prediction results. In this 3D-CNN, a linear regression model based on extracted features was used to predict the final property. A stochastic gradient descent (SGD) was employed to train the network. The learning rate was fixed to a constant at the beginning and was multiplied by 0.1 when the training loss did not improve. A 10-fold cross validation scheme was employed to avoid over-fitting and to quantify the errors.

Another CNN with different network structure and hyperparameters was applied to the prediction of the catalytic activity of catalytic Au nanoparticles from TEM images.²⁰¹ In this study, the input data is 2D images of atomic resolution STEM images instead of 3D microstructures. This CNN shares similar architectures with the one shown in Fig. 11a, except for the different dimension of input data and specialized filter size (3×3) due to the different characteristic scales in images. The aim of this study was to determine the presence of Au nanoparticle twins in the experimentally obtained HAADF-STEM images, which is a typical image classification issue.

There is broad scope for the adjustment of CNN structure for different applications where the microstructure is input in the form of images. The design choices include the number and



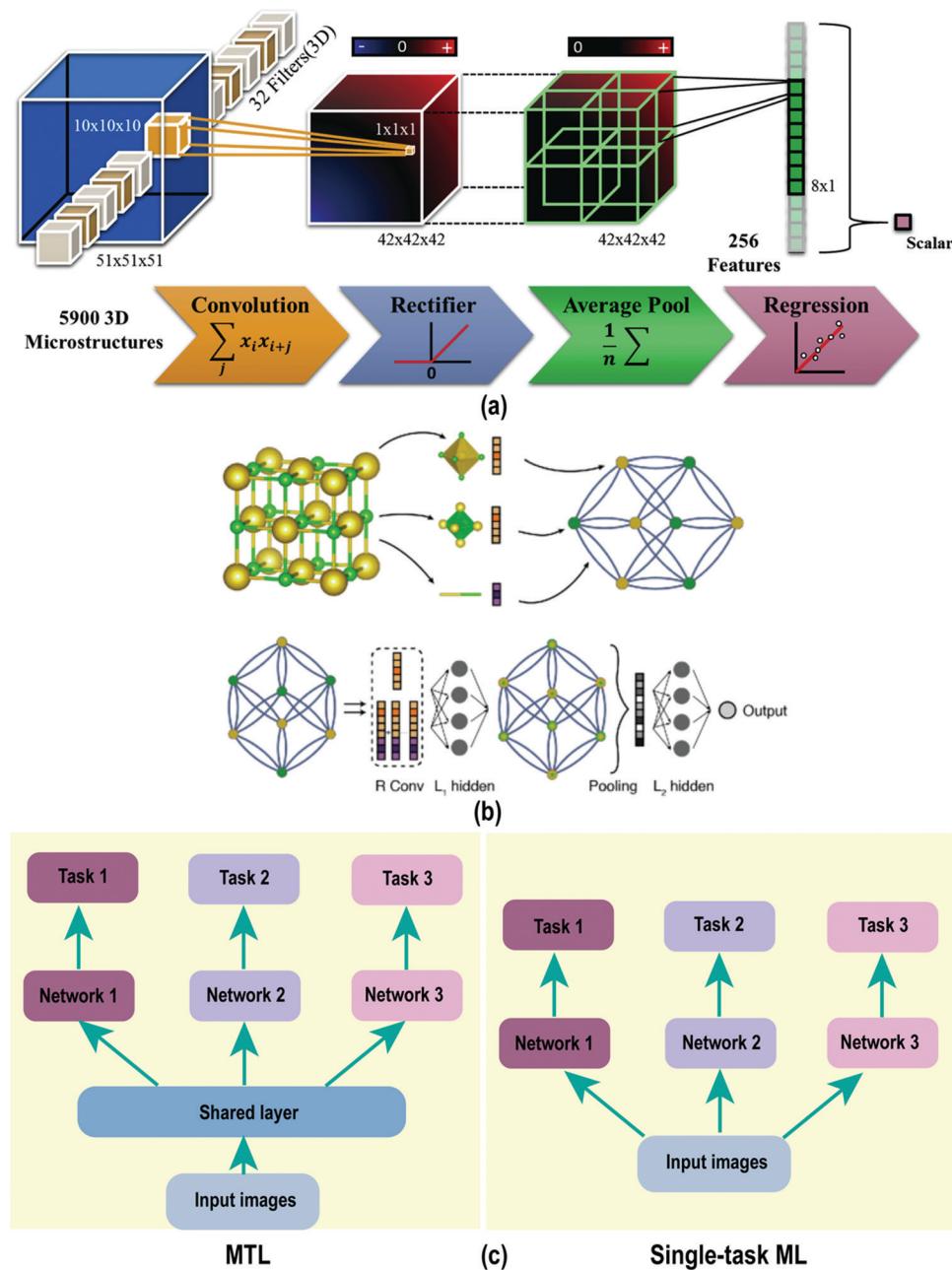


Fig. 11 (a) Description of a 3D-CNN for property prediction for microstructures¹⁸² (reproduced with permission from Elsevier); (b) illustration of the crystal graph convolutional neural network: construction of crystal graph and integration of the CNN with the crystal graph representation²⁰⁵ (reproduced with permission from American Physical Society); (c) architectures of MTL and single-task ML.

size of filters in convolutional layers, various pooling operations and activation functions, and combinations of these operations; these must, however, be tailored to the particular system of study in order to successfully extract and correlate the relevant features of the input data to the desired predicted property. CNNs have been transferred to segment and characterize the FIB-EBSD images of Li-ion electrode grains,²⁰² detect defects in Li-ion battery electrodes,²⁰³ classify images of perovskites²⁰⁴ and evaluate catalytic activity of Au nanoparticles from TEM images.²⁰¹ However, CNNs need special treatment

when they are employed to study graph-based crystals where atoms and bonds are represented as nodes and edges, respectively.^{205,206} For example, Xie *et al.*²⁰⁵ demonstrated a crystal graph convolutional neural network framework (CGCNN) which directly learnt material properties from the connection of atoms in the crystal, providing a universal and interpretable representation of crystalline materials. The main idea of their method was to represent the crystal structure by a crystal graph that encodes both atomic information and bonding interactions between atoms, and then build a convolutional neural

network based on these graphs and automatically extract characteristics and predict target properties, shown in Fig. 11b. This CGCNN was later extended by other workers to include atomic orbital interactions.²⁰⁶ It is noted that CNNs are popular in energy material science and molecular property prediction because of their strong capability in feature extraction, which is highly efficient compared with other fully connected neural networks.

The design of optoelectronic materials relies on prediction and tuning of both their ground-state and excited-state properties. In a study of conjugated polymers, Jackson *et al.*²⁰⁷ utilized molecular dynamics combined with semi-empirical quantum mechanical simulations of the optoelectronic properties to train a neural network based on a coarse-grained representation, enabling access to larger time and length scales. The coarse-graining was achieved using feature extraction with a 1D CNN, with inputs in form of the atomic reciprocal distance matrices. The CNN used a 1×1 kernel and an exponential linear unit activation function, with a parameter initialization using the LeCun normal initializer. This reduced-dimensional representation feeds into a bidirectional long short-term memory network (LSTM) to compute ground-state and excited-state properties, *e.g.* the charge density distributions of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), excited-state energies and linear optical absorption spectra, amongst others. Their ML model predicted the HOMO-2 to LUMO+2 orbital energies with a cross-validated mean absolute error (MAE) of 19.2 ± 0.2 meV and a coefficient of determination of 0.992 ± 0.0001 . By employing ML methods to make predictions from coarse-grained representations, their model significantly accelerates the prediction of optoelectronic properties of conjugated polymer systems. Lu *et al.*²⁰⁸ employed four state-of-the-art DNNs to predict the optoelectronic properties of oligothiophenes (OTs) which are organic semiconductor materials being explored for use in a range of optoelectronic devices. The four DNNs differ in their molecular representations and structure; (a) a deep tensor neural network (DTNN), (b) SchNet, a similar representation but using a continuous CNN (that can work with unequally spaced data, rather than pixels) (c) a message-passing neural network (MPNN) and (d) a multilevel graph convolutional neural network (MGCNN) with a similar representation to the MPNN. These networks were trained with 80 000 OT configurations generated by an MD model and optoelectronic properties calculated by DFT and TD-DFT models. Their results indicated that SchNet gave the best performance in predicting the excited-state properties of OTs and achieved MAEs below 0.1 eV even with a dataset as small as 5000 structures.

Though CNNs have demonstrated remarkable performance in various tasks and are often key components within neural networks, their black-box nature is still unresolved and the internal decisions within CNNs have often been poorly understood. Thus, an enhanced interpretability is necessary for CNNs to ensure that neural networks do not ignore relevant patterns in the dataset and to understand the underlying reasons for their outputs, thereby enhancing the accuracy and reliability of neural networks. In this regard, explainable AI (XAI) has become a

popular branch of research to interpret the behaviour of CNNs.²⁰⁹ Zhang *et al.*²¹⁰ quantitatively explained the rationale for each prediction from a pre-trained CNN by using decision trees to identify the basis for the outputs produced by the CNN.

In general, where the prediction of multiple properties is required, separate neural networks are trained for each specific property. The same feature may therefore be repeatedly extracted in the individual networks, which greatly wastes computational resources. To tackle this issue, a multi-task learning (MTL) method has been proposed in which a network layer is shared across different tasks²¹¹ so that common features are extracted more efficiently, as shown in Fig. 11c. MTL is particularly suitable for predicting the multi-physics properties of porous energy materials. While there are few reported studies of MTL relating to porous energy materials so far, Wang *et al.*²¹² employed an FEM model and a data-driven MTL model to predict three viscoelastic properties of polymer nanocomposites, namely the $\tan \delta$ peak, glassy modulus and rubbery modulus. In this model, the MTL was achieved through hard parameter sharing. The input image is first fed into a series of shared convolution and pooling layers to extract the high level shared structural features for different tasks. Then, three sets of task-specific layers, including one more convolution and pooling layer and two fully connected (FC) layers, are applied to predict the different output properties.

4.3 Predicting the performance of energy materials in service

The structure and properties of porous energy materials are the fundamental factors that impact on their performance in service. To understand the effects of structure and properties on in-service performance of porous energy materials, numerical modelling^{213–220} and *in situ/operando* experimental testing techniques^{81–88} have been used for bottom-up design and optimisation. In experimental testing, in-depth data analysis is necessary to extract important features and parameters from raw data. In numerical modelling, reduction of timing-consuming iterations is desired. To tackle these issues, data-driven models based on AI techniques have been developed to achieve in-depth data mining and rapid prediction of in-service performance based on experimental and numerical databases. Data-driven models are essentially producing an empirical non-linear fit to the multi-dimensional dataset that traditionally derives from experimental data but sometimes could be simulation data. In a typical data-driven model, the training data includes the operating conditions, the spatial distribution of physical and performance descriptors of energy devices, such as temperature, pressure, humidity, efficiency, output voltage and so on. These computational or experimental data are then fed into the data-driven machine-learning model for training. The selection of model can be diverse, depending on the particular requirements. For energy devices, ANN, SVM and others have been used. When the trained neural network has been validated with a test set of data, it can be put forward to predict information of interest such as overall performance or the spatial distribution of physical fields such as temperature and potential. Fig. 12a shows the workflow of a typical AI-based data-driven model. Wang *et al.*^{176,221} integrated two ML algorithms, *i.e.* SVM



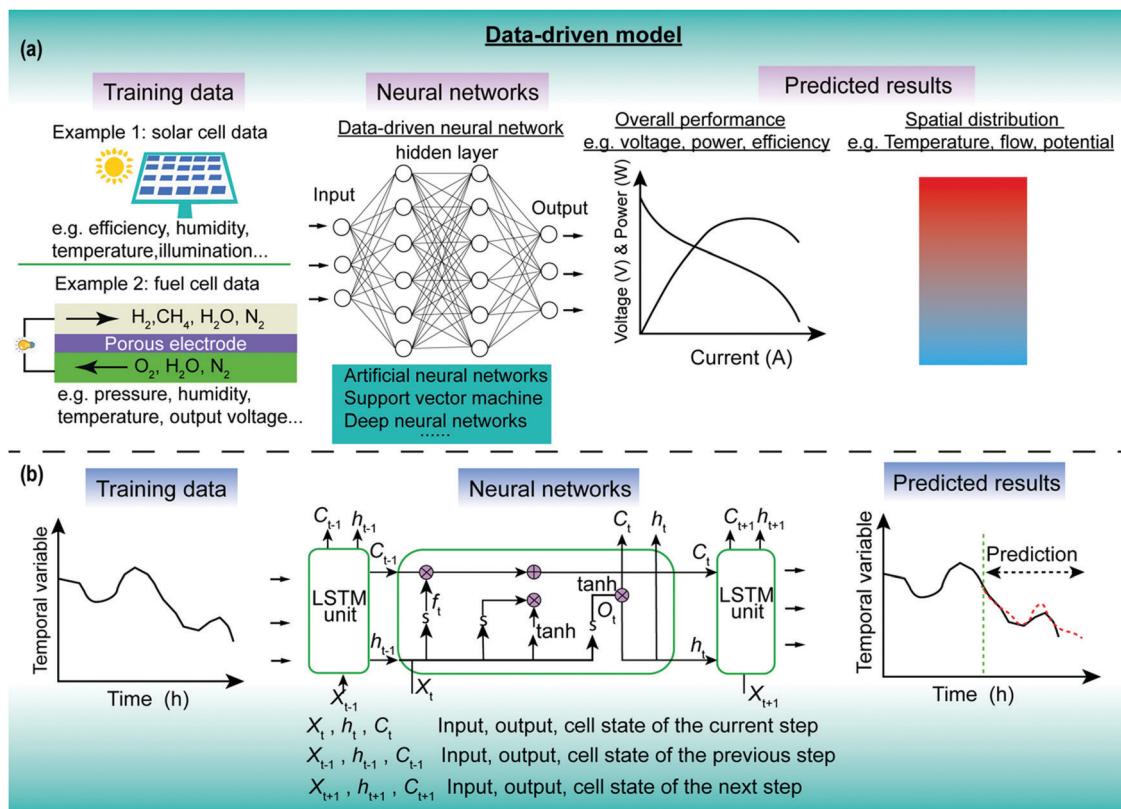


Fig. 12 (a) Workflow of typical data-driven neural networks for predicting the performance of energy devices including porous energy materials. (a) Steady-state performance;²²⁸ (b) transient performance.²²⁹

and ANN, with a 3D PEMFC multi-physics-resolved model. When the ML frameworks were well trained, this data-driven approach predicted cell performance within 1s, achieving a substantial computational acceleration compared with the conventional a 3D PEMFC model (~ten minutes for a single cell model). Xu *et al.*²²² further combined a DNN with a multi-physics SOFC model to optimise cell performance. The training sets in the data-driven models of Wang *et al.*^{176,221} and Xu *et al.*²²² are from numerical models of the system performance, which were validated with experimental data. Although these training sets are easier to obtain and lower-cost than experimental data, the errors of numerical models caused by the assumptions and simplifications of the model will affect the neural network training.

Besides integrating with numerical databases, AI can also work on testing data from experiments and forecast the lifetime and health of energy materials.^{223–225} Howard *et al.*²²³ adopted an ANN to identify the optimum operating parameters for re-aprest–recovery cycle of perovskite photovoltaics. In this data-driven model, the ANN is validated by predicting the performance and operations of a large set of solar cells from different laboratories, which is critical in assessing the ANN's generality and accuracy. When collecting experimental data for AI training, care must be taken to collect sufficient data for input and ensure that the variance in the training data should be representative of the true variance arising from all possible sources such as

different fabrication facilities and instrumentation variation. AI-based data-driven models can also be used to predict the lifetime of Li-ion batteries. Wu *et al.*²¹³ were the first to present perspectives on the integration of state-of-the-art battery modelling, in-vehicle diagnostic tools, data driven modelling approaches and emerging ML methods towards creating a battery digital twin. Severson *et al.*²²⁴ developed a logistic regression model to classify batteries into either a low-lifetime or a high-lifetime group, using only the first five cycles for batteries with various cycle life thresholds (generally 150 to 2300 cycles). Their model was capable of predicting the overall lifetime of a battery from its voltage levels and other information from its first 100 cycles with a 91% accuracy. This data-driven algorithm is highly significant in reducing the cost of battery development. Although this data-driven model does not require prior knowledge of degradation mechanisms, it is not able to indicate the mechanism occurring in any system, and therefore other experimental tools are needed to identify this. In addition, the model must be retrained for different battery designs.

Temporal characteristics of energy devices containing porous energy materials are also crucial for their performance and durability. The neural networks for predicting spatiotemporal characteristics are different from those for steady state because they require knowledge of the previous state and current state to predict the next state. Recurrent neural networks (RNNs) have been proposed to deal with these time series problems.^{226–232}

Similar to the commonly used feedforward network, a traditional RNN consists of input layer, hidden layer and output layer. The difference between RNNs and feedforward networks is the self-recurrent connection of the neurons in RNNs, *i.e.* the input of RNN neurons is not just the current input, but also what they have obtained previously in time.²²⁷ RNNs use their internal memory to process sequences of inputs. However, classic RNNs poorly describe long-term dependencies due to the vanishing-gradient phenomenon, which prevents the utilisation of the popular stochastic gradient-descent method used in training processes.²²⁶ To address this problem, the long short-term memory (LSTM) recurrent neural network, one kind of RNN, has been popular for learning long-term sequences. The LSTM allows the gradient to flow unchanged by employing a cell memory and improves the gradient vanishing and exploding issues.^{227,228} Fig. 12b shows the workflow of a LSTM RNN predicting the temporal evolution of the performance of in-service porous energy materials. Xie *et al.*²²⁸ fused a particle filter (a common mathematical algorithm in signal processing) and the LSTM RNN to predict the lifetime of a PEMFC. This integrated structure showed reasonable prediction when the training phase was large (60% of the dataset). However, the prediction capability was poor when using fewer temporal sequences as training data (training phase 40%). Other popular applications of LSTM have also been achieved in predicting wind speed,²²⁹ turbulent flow,²³⁰ state of charging of Li-ion batteries²³¹ and power forecasting of solar photovoltaic facilities.²³² Notably, these purely data-driven neural networks directly learn from data without explicitly enforcing physical constraints, and the predicted results may therefore deviate from the real data due to the absence of physical meanings in these neural networks. To further improve the prediction capability, physics-informed data-driven models have been on trend very recently.^{233,234} In these models, prior physical knowledge, such as pre-defined physical constraints, is incorporated into neural networks. Wang *et al.*²³³ developed a physics-informed DNN (TF-Net) which incorporated turbulent physics into a deep CNN to predict spatiotemporal turbulent flow. Their neural network shows higher prediction accuracy than other purely data-driven networks. Although this physics-informed model is for predicting turbulent flow, the approach shows great potential for predicting other transient physical phenomena in porous energy materials and energy devices in a physically meaningful way.

As mentioned in Section 3.1.3, higher dimensional imaging techniques (4D, 5D or generally denoted nD here) can quantitatively reveal vital physico-chemical dynamics in spatial and temporal scales simultaneously, thereby providing fuller insight into the screening of energy material components, the tuning of material properties and optimisation of operating parameters. To the best knowledge of authors, no nD imaging works have been integrated with AI techniques so far. Here, we propose two promising areas where AI techniques can greatly enhance the development of energy materials science based on nD imaging techniques: (1) discovering complex features from high-dimensional data. The data included in nD images are usually high-dimensional (*e.g.* number of datapoints in each 3D

image) and it is therefore challenging to reveal the relationships between many physical properties in this high-dimensional parameter space. In this regard, DNNs are powerful to extract correlations among parameters at different dimensional scales. For example, the correlation of catalyst activity with spatiotemporal microstructure evolution can be extracted by using a hybrid model of CNNs and LSTM to analyse 5D images (3D catalyst layer + temporal evolution + product water distribution) of *operando* PEM fuel cells; (2) forecasting long-term dynamics. Although long time-scale 5D imaging techniques can provide valuable insight into the durability and degradation of energy materials, the high cost of the facility requirements for 5D imaging hinders its advancing use. Thus, it is beneficial to use measured data collected over relatively short time scales to predict the long-term dynamics of the system. Moreover, it would be even more pertinent to predict long-term dynamics directly from structural and spatial chemical species distributions. In this regard, spatiotemporal neural networks like LSTM can forecast long-term 5D dynamics in energy materials based on finite 5D imaging data from experiments, thereby greatly reducing the cost of research on durability and degradation mechanisms. The work from Severson *et al.*²²⁴ and Xie *et al.*²²⁸ which employed ML methods to predict battery life and fuel cell dynamics, respectively, are worthy of reference in this regard. Furthermore, the use of techniques such as GANNs, have proven to be effective in generating large synthetic microstructures from sparse datasets and also in generating higher dimensional data from lower order information (*e.g.* 3D reconstructions from 2D data).¹⁸⁹ Thus, a promising route to address the potentially limited availability of nD datasets is to apply these GANN techniques towards training networks that are able to predict their nD properties without having to measure them.

4.4 New materials discovery and design

Generally, new materials can be achieved by composition modification and microstructural design. High throughput screening is often unavoidable for discovery of optimal compositions due to the vast search space, and this is usually time-consuming.^{235,236} Microstructural design often needs to adjust the operating parameters of manufacturing processes, which involves potentially numerous physical and chemical processes. Recently, the combination of energy material digitalisation and AI techniques opens opportunities to accelerate materials discovery.

In a digital-led high throughput screening process, the modelling techniques presented in Section 3 (*e.g.* DFT) provide training data for data-driven models. These data-driven models rapidly predict the material properties of a portfolio of candidates.^{237–241} Ma *et al.*²⁴¹ successfully discovered several potential two-dimensional optoelectronic octahedral oxyhalides with satisfactory properties (band gaps, high electron mobilities and ultrahigh absorbance coefficients) by integrating gradient boosted regression (GBR) with a DFT model. They trained the GBR model with 300 two-dimensional octahedral oxyhalides whose properties were simulated by DFT and subsequently screening another 5000 candidates. Aside from using data-driven models to directly predict material properties, AI techniques can also



accelerate material screening by extracting material features from experimental images (e.g. XRD images).^{242,243} In an experimental screening of perovskite-inspired materials, Sun *et al.* adopted a deep neural network to assist with classification and structural characterisation from XRD images.²⁴² This neural network can also diagnose the dimensionality of novel materials. With the help of AI techniques, an acceleration of over an order of magnitude per experimental learning cycle was achieved. One of the limitations of these high-throughput screening programmes is that of poor autonomy. Recently, exciting AI-based robots that run autonomously have been designed to perform this difficult experimental screening task.^{244,245} These AI-based robots successfully screen high-performance photocatalysts and antisolvents from several hundred candidates in several days. Also, an AI-based closed-loop optimisation of fast-charging protocols for batteries was proposed by Attia *et al.*²⁴⁶ This methodology autonomously incorporates feedback from past experiments to inform future decisions, which is highly meaningful for design of time-intensive experiments and multi-dimensional design spaces.

In microstructural design, few works have focussed on the optimisation of microstructures of porous energy materials in combination with their manufacture.^{247–250} A digital twin of the manufacturing process is necessary at this stage to virtually predict the influence of the manufacturing parameters on the final microstructures of porous energy materials. Notably, stochastic grain stacking methods and mesoscopic physical models contribute to the development of a digital twin since they can output virtually realistic microstructures of porous energy materials. Compared with stochastic grain stacking methods, mesoscopic single-phase models have been suggested as preferred methods because they consider the physical dynamics of separate materials in the manufacturing processes of porous energy materials. For example, DEM can capture aspects of the manufacturing process and particle–particle interactions that are relevant in determining the characteristics of the microstructure. Takagishi *et al.*²⁴⁸ and Lombardo *et al.*²⁴⁹ employed a mesoscopic coarse grained molecular dynamics model to digitalise the manufacturing process of Li-ion battery electrodes which involves multi-scale materials such as slurry, consisting of active materials, carbon additives, binders and solvents. Consequently, data-driven models can be further introduced to predict the properties of microstructures generated in a manufacturing process under specific conditions.

Digital twins for the fabrication and performance of porous energy materials are still largely absent and therefore their development is fruitful prospect moving forward. More digital twins such as the electrode model for Li-ion batteries should be developed by adopting physically meaningful microstructure construction models. For example, a digital twin of microstructures of SOFC electrodes can be achieved by integrating a phase field model and AI-based data-driven model. Moreover, autonomous optimisation algorithms based on digital twins and data-driven approaches that have been employed in porous energy material design have not yet been fully exploited for the optimisation of the complete system from material manufacture through to device operation. Recently, deep reinforcement

learning (RL) algorithms have been successfully employed to optimise the design of molecules.²⁵¹ It is feasible to extend this RL algorithm to the optimisation of microstructures of porous energy materials by integrating a digital twin of the manufacturing process and data-driven prediction models.

4.5 The importance of data in AI

As already mentioned, ML models often require a large amount of high-quality data that is vital for the accuracy and reliability of models. In fields where AI is well-established, such as image classification, large datasets are readily available, and ML models can be designed with high dimensionality (many learnt parameters). These methods can also benefit from learning a reduced set of relevant features from the inputs through techniques such as regularisation or other feature reduction methods. However, in engineering applications, data is often costly and difficult to obtain and datasets are of necessity limited in size. Therefore, strategies are required to achieve successful application of ML in such a context; these focus both on the efficient collection of data and the design of ML methods to work with smaller datasets.

Currently, there are three common methods that have been adopted by researchers to prepare training data for their ML models: (1) direct experimental measurement; (2) simulation or (3) meta-analysis. Whilst direct experimental measurement is a straightforward approach, it can be difficult and expensive to access to facilities and obtain data. The measured data has associated experimental errors, and one significant problem is the challenge of fully characterising the material structure for which the property was measured, especially taking into account sample variability. Simulations using physically-based models offer an alternative source of data. Modelling tools such as those introduced in Section 3.3 such as DFT, MD, FVM have been widely employed to prepare training datasets for NNs to achieve high-throughput material structure screening as well as for rapid property and performance prediction. An alternative to these data collection techniques is to use a meta-analysis approach, in which experimental or simulation data is mined from literature and data repositories; some of these are “open” and are hosted by academic, non-profit or governmental communities. For example, numerous experimental testing data and 3D microstructures of Li-ion batteries have been released by the Battery Microstructure Project of ETH Zurich²⁵² and National Renewable Energy Laboratory of the USA.²⁵³ These high-quality data can be used as inputs for GANNs for electrode generation, and for NN systems tailored for the prediction of the performance and lifetime of Li-ion batteries. Yildirim *et al.*²⁵⁴ prepared a hysteresis dataset of perovskite solar cells by collecting data from 194 articles. With this dataset, they employed ML method to analyse the hysteresis and reproducibility of perovskite solar cells and their relations with power conversion efficiency and long-term stability. Furthermore, approaches such as text mining of the existing academic literature for key performance metrics associated with energy systems have been demonstrated. For example, Torayev *et al.*²⁵⁵ showcased a text mining system



that analysed 1800 publications relating to Li–O₂ batteries, extracting metrics such as the discharge capacity automatically.

Whilst experimental data can be costly and challenging to obtain, simulations of complex materials can also be resource-hungry and time-consuming. Collection of data by either means may therefore need to be carefully targeted in order to obtain the best outcome from network training with the minimum amount of data. Janet *et al.* adopted a multidimensional “expected improvement criterion” to estimate how the use of a proposed new piece of training data would improve the predictive accuracy of the system.²⁵⁶ They were therefore able to generate new data in the regions (of feature space) where that data would make the greatest improvement to the predictive power of the network. The investigation focussed on the discovery of candidate transition metal complexes acting as redox couples for redox flow batteries, and used resource-intensive DFT simulations to generate data; targeting the data collection enabled these to be used in a highly efficient way. An alternative way to address the data problem is to adopt a transfer learning approach, in which models already trained for one task are used as a building block for a different task. Thus, the hierarchy of features learned by the pretrained network can be used for many other different tasks with a much-reduced training demand.²⁵⁷ Badmos *et al.*²⁰³ achieved AI-based defect detection in Li-ion battery electrode based on limited images through use of the transfer learning method.

Having considered the challenges and potential strategies around data collection, we now address the question of effective ML design, particularly when working with limited data. The more complex the system, the more data is required for training the model. Therefore, an important strategy to reduce the data demand is to reduce the number of features (inputs) to the model. Broadly speaking, approaches to achieve this fall into two categories: feature engineering and feature selection.²⁵⁸ Feature engineering refers to the design of input features using transformations of the raw data; these features may be based on physical or engineering expertise to target relevant physical quantities, or may be extracted from the data using statistical measures *etc.* In their study of battery lifetime prediction,²²⁴ Severson *et al.* adopted features based on the differences between repeated charging cycles. The use of such a reduced set of engineered features may result in a highly predictive model but is at risk of introducing model bias. An alternative approach is that of feature selection or feature reduction in which the system learns which features are the most important and others can then be removed from the model. A wide range of techniques exist, including the widely used regularisation methods, filters (eliminating a subset of features) and wrapper methods (to test the elimination of features). Janet and Kulik²⁵⁸ explored feature selection by a range of techniques for the design of transition metal complexes and were able to identify the importance of short-range electronic properties (*e.g.* atomic numbers of atoms local to the metal atom, *etc.*) as well as longer-range steric effects. They also used feature engineering approaches to design an appropriate representation of the system as a set of features which were subsequently down-selected.

4.6 Perspectives

Based on the knowledge of porous energy materials reviewed in this paper, we propose a future route for autonomous optimisation and rational evaluation of microstructures of porous energy materials, as shown in Fig. 13. In this route, physical analysis techniques like XRD pass basic material information (*e.g.* composition), experimental testing data and material search boundaries to a deep reinforcement learning (RL) algorithm. The digital space includes a deep RL algorithm that integrates two models, *i.e.* generative and predictive models. The generative model (G-model) mainly consists of digital twins of manufacturing processes of porous energy materials. The G-model is able to output a large set of candidate digital microstructures that correspond to specified manufacturing parameters. Microstructures of porous energy materials generated by the G-model are transferred to the predictive model (P-model) that is specialised in assessing their properties and performance. This predictive model typically consists of multi-scale models, performance models and DNNs, where these DNNs are trained by the data from models. The role of the P-model is to act as a critic of the performance of the G-model by a numerical reward fed back to the generated microstructures. The G-model takes this reward into account and adjusts the manufacturing parameters, then outputs new microstructures of porous energy materials to maximise the expected reward. Finally, the down-selection performed by the deep RL reduces the large database into limited candidates which will be passed into the physical world again. Then, AI-based robots autonomously perform the rational evaluation by conducting checks on the feasibility of manufacture, *in situ* characterisation, durability tests and so on to determine the best microstructures and corresponding manufacturing parameters. If the rational evaluation is unsatisfactory, these testing data will be feedback both in the physical world, to adjust physical parameters, and also in the digital space to adjust the reward function of the RL algorithm to generate new candidates. It is noted that the physical world could also share the same AI techniques that have been employed for data processing in material digitalisation. For example, CNNs can be used to enhance the efficiency and accuracy of defect detection when assessing manufacturing feasibility. Data-driven modelling techniques can be used to predict the lifetime of energy materials based on experimental data on short time-scale durability, thereby substantially reducing the cost of the rational evaluation. This closed-loop optimisation roadmap may potentially reduce the load of AI-based robots and eventually accelerate the design and optimisation of porous energy materials. When the relevant data contains uncertainties, probabilistic ML methods such as Bayesian ML can be viable alternative approaches because they simplify the interpretability of the results and facilitate a subsequent optimisation process to find the optimum design for a given task.²⁵⁹

Furthermore, theoretical predictions of material properties will inherently have a degree of uncertainty. Thus, high-throughput synthesis techniques which combine with ML techniques to aid with the design of experiments will be essential. For instance,



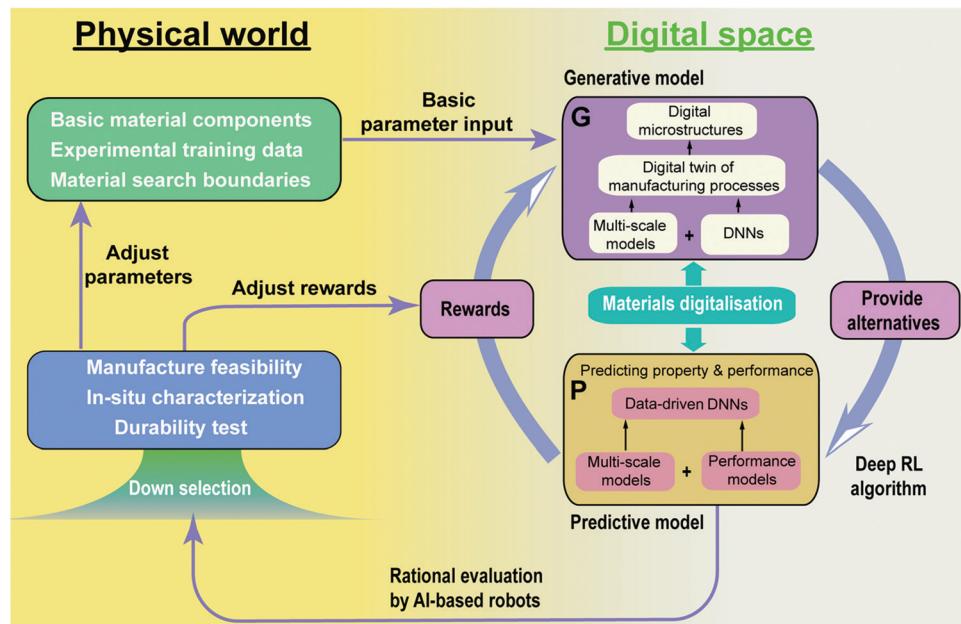


Fig. 13 A future potential route for the autonomous optimisation of microstructures based on digitalisation techniques of porous energy materials.

Granda *et al.*²⁶⁰ demonstrated how an ML-driven decision making system for an automated synthesis process with real time NMR and IR-spectroscopy was able to predict the reactivity of 1000 reaction combinations with an 80% accuracy. Thus, ML-enabled cyber-physical systems are a key enabler to bridge the uncertainty in computational predictions and practical energy materials.

5. Conclusions

Porous energy materials have served society for many centuries. The understanding of porous energy materials showed unique characteristics in different eras depending on the level of theory and technology at the time. In spite of the boom in innovative materials and energy technologies, the design and optimisation of their structures and operating parameters remains challenging and requires further effort from both industrial and academic communities. Thus, shaping all materials into virtual data spaces instead of relying on experimental investigation has gradually attracted increasing attention. This appealing target was recently promoted by the surge in artificial intelligence (AI) in material science that has revolutionised the way we model the structures and properties of porous energy materials. Based on the digital porous energy material spaces created by the integration of AI techniques, conventional theories and technologies, we anticipate that faster discovery, design and optimisation of porous energy materials will be readily achieved. We have stepped into a society of the internet of things, where digital transformation of every element and linked chains is essential to create efficient end-to-end information streams. Of course, as core components of energy devices, porous energy materials need to be fully digitised from their microstructure to their properties and through to in-service performance, extending to their discovery, design and optimisation. We have proposed a future roadmap for an

automated optimisation route for microstructures based on digitalisation techniques of porous energy materials, deep reinforced learning methods and AI-informed robotic automation. We envisage that this closed-loop optimisation roadmap will accelerate the search for optimised porous energy materials, will lead to the discovery of potentially novel structures and will effectively target the load on AI-based robots performing the final physical material screening. In the physical world of this roadmap, high-throughput synthesis techniques based on cyber-physical systems will be essential.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

The authors would like to acknowledge the support of the Royal Society – K. C. Wong International Fellowship (NIF/R1/191864) from the Royal Society, the UK-China International Cooperation and Exchange Project (Newton Advanced Fellowship) jointly supported by the National Natural Science Foundation of China (51861130359) and the UK Royal Society (NAF/R1/180146), the Faraday Institution Multi-Scale Modelling project ((EP/S003053/1, grant number FIRG003), and the support from Engineering and Physical Sciences Research Council, UK (EP/S000933/1 and EP/V011863/1).

References

- 1 W. Li, J. Liu and D. Zhao, *Nat. Rev. Mater.*, 2016, **1**, 1–17.
- 2 M. R. Lukatskaya, B. Dunn and Y. Gogotsi, *Nat. Commun.*, 2016, **7**, 1–13.



3 A. Alhosani, A. Scanziani, Q. Lin, A. Raeini, B. Bijeljic and M. Blunt, *Sci. Rep.*, 2020, **10**, 8534.

4 P. Pietsch, D. Westhoff, J. Feinauer, J. Eller, F. Marone, M. Stampanoni, V. Schmidt and V. Wood, *Nat. Commun.*, 2016, **7**, 1–11.

5 Z. Chen, P. Li, R. Anderson, X. Wang, X. Zhang, L. Robison, L. R. Redfern, S. Moribe, T. Islamoglu and D. A. Gómez-Gualdrón, *Science*, 2020, **368**, 297–303.

6 J. Lai, G. Wang, Z. Wang, J. Chen, X. Pang, S. Wang, Z. Zhou, Z. He, Z. Qin and X. Fan, *Earth-Sci. Rev.*, 2018, **177**, 436–457.

7 T. S. Pathan, M. Rashid, M. Walker, W. Widanage and E. Kendrick, *J. Phys. Energy*, 2019, **1**, 44003.

8 N. Baharuddin, N. Rahman, H. Rahman, M. Somalu, M. Azmi and J. Raharjo, *Int. J. Energy Res.*, 2020, **44**(11), 8296–8313.

9 Y. Awin and N. Dukhan, *Appl. Energy*, 2019, **252**, 113458.

10 M. Sahoo and P. Kale, *Microporous Mesoporous Mater.*, 2019, **289**, 109619.

11 M. Liu, L. Wang, K. Zhao, S. Shi, Q. Shao, L. Zhang, X. Sun, Y. Zhao and J. Zhang, *Energy Environ. Sci.*, 2019, **10**, 2890–2923.

12 C. Stavrakas, A. Zhumekenov, R. Brenes, M. Abdi-jalebi, V. Bulovic, O. Bakr, E. Barnard and S. Stranks, *Energy Environ. Sci.*, 2018, **11**, 2846–2852.

13 Y. Nomura, K. Yamamoto, M. Fujii, T. Hirayama, E. Igaki and K. Saitoh, *Nat. Commun.*, 2020, **11**, 2824.

14 Y. Meng and M. Arroyo-de Dompablo, *Energy Environ. Sci.*, 2009, **2**, 589–609.

15 E. Ryan and P. Mukherjee, *Prog. Energy Combust. Sci.*, 2019, **71**, 118–142.

16 P. Mukherjee, Q. Kang and C. Wang, *Energy Environ. Sci.*, 2011, **4**, 346–369.

17 L. Kahle, A. Marcolongo and N. Marzari, *Energy Environ. Sci.*, 2020, **13**, 928–948.

18 Materials Genome Initiative, <https://www.mgi.gov/>, (accessed October 2020).

19 The Material Digitalization Platform, <https://www.material-digital.de/>, (accessed October 2020).

20 A. White, *MRS Bull.*, 2012, **37**, 715–716.

21 A. Dima, S. Bhaskarla, C. Becker, M. Brady, C. Campbell, P. Dessaix, R. Hanisch, U. Kattner, K. Kroenlein and M. Newrock, *JOM*, 2016, **68**, 2053–2064.

22 J. de Pablo, N. Jackson, M. Webb, L. Chen, J. Moore, D. Morgan, R. Jacobs, T. Pollock, D. Schlam and E. Toberer, *npj Comput. Mater.*, 2019, **5**, 41.

23 C. Cremer and B. Masters, *Eur. Phys. J. H*, 2013, **38**, 281–344.

24 K. Nakano, A. Ohbuchi, S. Mito and Z. Xue, *Energy Procedia*, 2014, **63**, 5754–5761.

25 R. Kandiyoti, A. Herod and K. Bartle, *Solid fuels and heavy hydrocarbon liquids: thermal characterization and analysis*, Elsevier, 2006, pp. 13–35.

26 N. Johnson, The legacy and future of CFD at Los Alamos, Los Alamos National Lab., NM (United States), 1996.

27 G. Lu, G. M. Lu and Z. Xiao, *J. Porous Mater.*, 1999, **6**, 359–368.

28 S. Durucan and J. Edwards, *Min. Sci. Technol.*, 1986, **3**, 205–216.

29 S. Shin, R. Baldwin and R. Miller, *Energy Fuels*, 1987, **1**, 377–380.

30 M. Leverett, *Trans. AIME*, 1941, **142**, 152–169.

31 C. Granqvist, *Adv. Mater.*, 2003, **15**, 1789–1803.

32 Y. Sun, Q. Wu and G. Shi, *Energy Environ. Sci.*, 2011, **4**, 1113–1132.

33 X. Han, Q. Wang, Y. Park, C. Tjoen, A. Sommers and A. Jacobi, *Heat Transfer Eng.*, 2012, **33**, 991–1009.

34 P. Kumar, J. Sundaramurthy, S. Sundarajan, V. Babu, G. Singh, S. Allakhverdiev and S. Ramakrishna, *Energy Environ. Sci.*, 2014, **7**, 3192–3222.

35 Z. Ma, X. Yuan, L. Li, Z. Ma, D. Wilkinson, L. Zhang and J. Zhang, *Energy Environ. Sci.*, 2015, **8**, 2144–2198.

36 X. Zhang, X. Cheng and Q. Zhang, *J. Energy Chem.*, 2016, **25**, 967–984.

37 F. Wang, X. Wu, C. Li, Y. Zhu, L. Fu, Y. Wu and X. Liu, *Energy Environ. Sci.*, 2016, **9**, 3570–3611.

38 H. Zhang, J. Nai, L. Yu and X. W. D. Lou, *Joule*, 2017, **1**, 77–107.

39 M. Mujeebu, M. Abdullah, M. Bakar, A. Mohamad, R. Muhamad and M. Abdullah, *J. Environ. Manage.*, 2009, **90**, 2287–2312.

40 L. Xu, S. Tang, Y. Cheng, K. Wang, J. Liang, C. Liu, Y.-C. Cao, F. Wei and L. Mai, *Joule*, 2018, **2**, 1991–2015.

41 J. Kim, S. Chai, Y. Ji, B. Levy-Wendt, S. Kim, Y. Yi, T. Heinz, J. Nørskov, J. Park and X. Zheng, *Adv. Energy Mater.*, 2018, **8**, 1801717.

42 W. Kan, A. Samson and V. Thangadurai, *J. Mater. Chem. A*, 2016, **4**, 17913–17932.

43 C. Chen, Y. Kuang and L. Hu, *Joule*, 2019, **3**, 683–718.

44 X. Wu, L. Peng, S. Xiang and W. Cai, *Phys. Chem. Chem. Phys.*, 2018, **20**, 30150–30158.

45 B. Li, H. Wen, W. Zhou and B. Chen, *J. Phys. Chem. Lett.*, 2014, **5**, 3468–3479.

46 F. Huang, A. Pascoe, W. Wu, Z. Ku, Y. Peng, J. Zhong, R. Caruso and Y. Cheng, *Adv. Mater.*, 2017, **29**, 1601715.

47 H. Ghasemi, G. Ni, A. Marconnet, J. Loomis, S. Yerci, N. Miljkovic and G. Chen, *Nat. Commun.*, 2014, **5**, 4449.

48 J. Zhao, M. Cano, J. Giner-Casares, R. Luque and G. Xu, *Energy Environ. Sci.*, 2020, **13**, 2618–2656.

49 M. Paul and B. Gates, *Sci. Rep.*, 2019, **9**, 1–9.

50 W. An, B. Gao, S. Mei, B. Xiang, J. Fu, L. Wang, Q. Zhang, P. K. Chu and K. Huo, *Nat. Commun.*, 2019, **10**, 1–11.

51 D. de Quilettes, S. Vorpahl, S. Stranks, H. Nagaoka, G. Eperon, M. Ziffer, H. Snaith and D. Ginger, *Science*, 2015, **348**, 683–686.

52 S. Chiodini, D. Reinares-Fisac, F. Espinosa, E. Gutiérrez-Puebla, A. Monge, F. Gádara and R. García, *Sci. Rep.*, 2017, **7**, 1–6.

53 Y. Yang, X. Zhang, L. Qin, Q. Zeng, X. Qiu and R. Huang, *Nat. Commun.*, 2017, **8**, 1–10.

54 D. Yadav, V. Ganesan, P. Sonkar, R. Gupta and P. Rastogi, *Electrochim. Acta*, 2016, **200**, 276–282.

55 A. Star and T. Fuller, *J. Electrochem. Soc.*, 2017, **164**, F901.



56 T. Terao, G. Inoue, M. Kawase, N. Kubo, M. Yamaguchi, K. Yokoyama, T. Tokunaga, K. Shinohara, Y. Hara and T. Hara, *J. Power Sources*, 2017, **347**, 108–113.

57 M. Kishimoto, Y. Kawakami, Y. Otani, H. Iwai and H. Yoshida, *Scr. Mater.*, 2017, **140**, 5–8.

58 B. Song, T. Sui, S. Ying, L. Li, L. Lu and A. M. Korsunsky, *J. Mater. Chem. A*, 2015, **3**, 18171–18179.

59 A. Cocco, G. Nelson, W. Harris, A. Nakajo, T. Myles, A. Kiss, J. Lombardo and W. Chiu, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16377–16407.

60 S. Kelly, H. El-Sobky, C. Torres-Verdín and M. Balhoff, *Adv. Water Resour.*, 2016, **95**, 302–316.

61 I. Zenyuk, D. Parkinson, L. Connolly and A. Weber, *J. Power Sources*, 2016, **328**, 364–376.

62 S. Normile and I. Zenyuk, *Solid State Ionics*, 2019, **335**, 38–46.

63 D. Kennouche, Y. Chen-Wiegart, C. Riscoe, J. Wang and S. Barnett, *J. Power Sources*, 2016, **307**, 604–612.

64 C. Rahe, S. Kelly, M. Rad, D. Sauer, J. Mayer and E. Figgemeier, *J. Power Sources*, 2019, **433**, 126631.

65 C. Zhao, T. Wada, V. De Andrade, D. Gürsoy, H. Kato and Y. Chen-Wiegart, *Nano Energy*, 2018, **52**, 381–390.

66 H. Zhao, T. Zhao, Z. Ning, R. Zhang, T. Duan, Q. Wang, P. Lian, D. Zhang and W. Zhang, *Mar. Petrol. Geol.*, 2019, **102**, 604–614.

67 H. Chang, A. Ilott, N. Trease, M. Mohammadi, A. Jerschow and C. Grey, *J. Am. Chem. Soc.*, 2015, **137**, 15209–15216.

68 A. Ilott, M. Mohammadi, C. Schauerman, M. Ganter and A. Jerschow, *Nat. Commun.*, 2018, **9**, 1–7.

69 S. Chandrashekhar, N. Trease, H. Chang, L. Du, C. Grey and A. Jerschow, *Nat. Mater.*, 2012, **11**, 311–315.

70 W. Rose, H. Haas, A. Chen, N. Jeon and L. Lauhon, *Phys. Rev. X*, 2018, **8**, 011030.

71 B. Moores, A. Eichler, Y. Tao, H. Takahashi, P. Navaretti and C. Degen, *Appl. Phys. Lett.*, 2015, **106**, 213101.

72 R. Stannarius, *Rev. Sci. Instrum.*, 2017, **88**, 051806.

73 H. Wang, T. Koster, N. Trease, J. Segalini, P. Taberna, P. Simon, Y. Gogotsi and C. Grey, *J. Am. Chem. Soc.*, 2011, **133**(48), 19270–19273.

74 O. Pecher, D. Halat, J. Lee, Z. Liu, K. Griffith, M. Braun and C. Grey, *J. Magnet. Reason.*, 2017, **275**, 127–136.

75 X. Li, X. Wei and Q. Chen, *Int. J. Nanomanuf.*, 2016, **12**, 264–277.

76 N. Shpigel, M. Levi, X. Cheng, T. Cao, R. Wu, T. Mathis, Y. Zhang, D. Aurbach and Y. Gogotsi, *ACS Energy Lett.*, 2019, **4**(8), 1907–1917.

77 J. Nam, K. Lee, G. Hwang, C. Kim and M. Kaviany, *Int. J. Heat Mass Transfer*, 2009, **52**, 2779–2791.

78 A. Kardjilov, I. Manke, A. Hilger, M. Strobl and J. Banhart, *Materials Today*, 2011, **14**(6), 248–256.

79 P. Boillat, E. Lehmann, P. Trtik and M. Cochet, *Curr. Opin. Electrochem.*, 2017, **5**, 3–10.

80 Y. Wu, J. Cho, M. Whiteley, L. Rasha, T. Neville and R. Ziesche, *et al.*, *Int. J. Hydrogen Energy*, 2020, **45**(3), 2195–2205.

81 J. Paz-Garcia, O. Taiwo, E. Tudisco, D. Finegan, P. Shearing, D. Brett and S. Hall, *J. Power Sources*, 2016, **320**, 196–203.

82 R. White, A. Wu, M. Najm, F. Orfino, M. Dutta and E. Kjeang, *J. Power Sources*, 2017, **350**, 94–102.

83 Y. Nagai, J. Eller, T. Hatanaka, S. Yamaguchi, S. Kato, A. Kato, F. Marone, H. Xu and F. N. Büchi, *J. Power Sources*, 2019, **435**, 226809.

84 J. Wu, L. G. Melo, X. Zhu, M. West, V. Berejnov, D. Susac, J. Stumper and A. Hitchcock, *J. Power Sources*, 2018, **381**, 72–83.

85 T. Heenan, D. Finegan, B. Tjaden, X. Lu, F. Iacoviello, J. Millichamp, D. Brett and P. Shearing, *Nano Energy*, 2018, **47**, 556–565.

86 T. Bultreys, M. Boone, M. Boone, T. De Schryver, B. Masschaele, L. Van Hoorebeke and V. Cnudde, *Adv. Water Resour.*, 2016, **95**, 341–351.

87 A. Macente, F. Fusseis, I. Butler, E. Tudisco, S. Hall and E. Andò, *Earth Planet. Sci. Lett.*, 2018, **502**, 115–125.

88 H. Menke, C. Reynolds, M. Andrew, J. Nunes, B. Bijeljic and M. Blunt, *Chem. Geol.*, 2018, **481**, 27–37.

89 D. Finegan, A. Vamvakarios, L. Cao, C. Tan, T. Heenan, S. Daemi, S. Jacques, A. Beale, M. Michiel and K. Smith, *Nano Lett.*, 2019, **19**, 3811–3820.

90 T. Li, T. Heenan, M. Rabuni, B. Wang, N. Farandos, G. Kelsall, D. Matras, C. Tan, X. Lu and S. Jacques, *Nat. Commun.*, 2019, **10**, 1–11.

91 Ultra-high resolution Resolution Scanning Electron Microscope SU9000, <https://www.hitachi-hightech.com/global/science/products/microscopes/electron-microscope/fe-sem/su9000.html> (accessed October 2020).

92 J. Woods, N. Bhattacharai, P. Chapagain, Y. Yang and S. Neupane, *Nano Energy*, 2019, **56**, 619–640.

93 A. Quinn, H. Moutinho, F. Usseglio-Viretta, A. Verma, K. Smith, M. Keyser and D. Finegan, *Cell Rep. Phys. Sci.*, 2020, **1**, 100137.

94 M. Holz, C. Reuter, A. Reum, A. Ahmad, M. Hofmann, T. Ivanov, S. Mechold and I. W. Rangelow, *Microelectron. Eng.*, 2020, **224**, 111234.

95 A. Cattoni, D. Mailly, O. Dalstein, M. Faustini, G. Seniutinas, B. Rösner and C. David, *Microelectron. Eng.*, 2018, **193**, 18–22.

96 W. Van Aarle, W. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabrowski, J. De Beenhouwer, K. Batenburg and J. Sijbers, *Opt. Express*, 2016, **24**, 25129–25147.

97 P. Pietsch and V. Wood, *Annu. Rev. Mater. Res.*, 2017, **47**, 451–479.

98 J. Hack, P. García-Salaberri, M. Kok, R. Jervis, P. Shearing, N. Brandon and D. Brett, *J. Electrochem. Soc.*, 2020, **167**, 013545.

99 N. Michailidis, F. Stergioudi, H. Omar, D. Papadopoulos and D. Tsipas, *Colloids Surf. A*, 2011, **382**, 124–131.

100 H. Pourmodheji, E. Ghafar-Zadeh and S. Magierowski, *Sensors*, 2016, **16**, 850.

101 R. Bhattacharyya, B. Key, H. Chen, A. Best, A. Hollenkamp and C. Grey, *Nat. Mater.*, 2010, **9**, 504–510.

102 S. Daemi, C. Tan, A. Vamvakarios, T. M. Heenan, D. Finegan, M. Di Michiel, A. M. Beale, J. Cookson, E. Petrucco and J. Weaving, *Phys. Chem. Chem. Phys.*, 2020, **22**, 17814–17823.



103 W. Zhang, Y. Tian, H. he, L. Xu and D. Zhao, *Natl. Sci. Rev.*, 2020, **7**(11), 1702–1725.

104 S. Nong, W. Dong, J. Yin, B. Dong, Y. Lu, X. Yuan, X. Wang, K. Bu, M. Chen, S. Jiang, L. Liu, M. Sui and F. Huang, *J. Am. Chem. Soc.*, 2018, **140**(17), 5719–5727.

105 A. Ilott, M. Mohammadi, H. Chang, C. Grey and A. Jerschow, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 10779–10784.

106 Q. Cheng, L. Wei, Z. Liu, N. Ni, Z. Sang, B. Zhu, W. Xu, M. Chen, Y. Miao and L. Chen, *Nat. Commun.*, 2018, **9**, 1–10.

107 A. Vamvakarios, S. Jacques, M. Di Michiel, D. Matras, V. Middelkoop, I. Ismagilov, E. Matus, V. Kuznetsov, J. Drnec and P. Senecal, *Nat. Commun.*, 2018, **9**, 1–11.

108 V. Schulz, J. Becker, A. Wiegmann, P. Mukherjee and C. Wang, *J. Electrochem. Soc.*, 2007, **154**, B419.

109 L. Hao and P. Cheng, *Int. J. Heat Mass Transfer*, 2012, **55**, 133–139.

110 Z. Niu, Z. Bao, J. Wu, Y. Wang and K. Jiao, *Appl. Energy*, 2018, **232**, 443–450.

111 J. Hinebaugh, J. Gostick and A. Bazylak, *Int. J. Hydrogen Energy*, 2017, **42**, 15872–15886.

112 V. Burganos, E. Skouras and A. Kalarakis, *J. Power Sources*, 2017, **365**, 179–189.

113 M. El Hannach, R. Singh, N. Djilali and E. Kjeang, *J. Power Sources*, 2015, **282**, 58–64.

114 Y. Nishida and S. Itoh, *Electrochim. Acta*, 2011, **56**, 2792–2800.

115 M. Neumann, J. Staněk, O. M. Pecho, L. Holzer, V. Beneš and V. Schmidt, *Comput. Mater. Sci.*, 2016, **118**, 353–364.

116 H. Moussaoui, J. Laurencin, Y. Gavet, G. Delette, M. Hubert, P. Cloetens, T. Le Bihan and J. Debayle, *Comput. Mater. Sci.*, 2018, **143**, 262–276.

117 H. Moussaoui, R. Sharma, J. Debayle, Y. Gavet, G. Delette and J. Laurencin, *J. Power Sources*, 2019, **412**, 736–748.

118 R. Thiedmann, O. Stenzel, A. Spettl, P. R. Shearing, S. J. Harris, N. P. Brandon and V. Schmidt, *Comput. Mater. Sci.*, 2011, **50**, 3365–3376.

119 W. Mai, M. Yang and S. Soghrazi, *Electrochim. Acta*, 2019, **294**, 192–209.

120 J. Feinauer, T. Brereton, A. Spettl, M. Weber, I. Manke and V. Schmidt, *Comput. Mater. Sci.*, 2015, **109**, 137–146.

121 S. Hein, J. Feinauer, D. Westhoff, I. Manke, V. Schmidt and A. Latz, *J. Power Sources*, 2016, **336**, 161–171.

122 D. Westhoff, J. Feinauer, K. Kuchler, T. Mitsch, I. Manke, S. Hein, A. Latz and V. Schmidt, *Comput. Mater. Sci.*, 2017, **126**, 453–467.

123 K. Kuchler, D. Westhoff, J. Feinauer, T. Mitsch, I. Manke and V. Schmidt, *Modell. Simul. Mater. Sci. Eng.*, 2018, **26**, 035005.

124 S. Kim and H. Pitsch, *J. Electrochem. Soc.*, 2009, **156**, B673.

125 W. Wu and F. Jiang, *Int. J. Hydrogen Energy*, 2014, **39**, 15894–15906.

126 W. Wu and F. Jiang, *Mater. Charact.*, 2013, **80**, 62–68.

127 S. He, J. Zeng, B. T. Habte and F. Jiang, *Sci. Bull.*, 2016, **61**, 656–664.

128 S. He, B. T. Habte and F. Jiang, *Solid State Ionics*, 2016, **296**, 146–153.

129 B. Habte and F. Jiang, *Solid State Ionics*, 2018, **314**, 81–91.

130 B. Habte and F. Jiang, *Microporous Mesoporous Mater.*, 2018, **268**, 69–76.

131 P. Yin and G.-F. Zhao, *Int. J. Rock Mech. Min. Sci.*, 1997, **2014**(70), 82–89.

132 H. Okabe and M. J. Blunt, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2004, **70**, 066135.

133 H. Okabe and M. J. Blunt, *J. Petrol. Sci. Eng.*, 2005, **46**, 121–137.

134 A. Hajizadeh, A. Safekordi and F. A. Farhadpour, *Adv. Water Resour.*, 2011, **34**, 1256–1267.

135 Y. Wu, C. Lin, L. Ren, W. Yan, S. An, B. Chen, Y. Wang, X. Zhang, C. You and Y. Zhang, *J. Nat. Gas Sci. Eng.*, 2018, **51**, 129–140.

136 M. Daino and S. Kandlikar, *Int. J. Hydrogen Energy*, 2012, **37**, 5180–5189.

137 D. Bohra, J. Chaudhry, T. Burdyny, E. Pidko and W. Smith, *Energy Environ. Sci.*, 2019, **12**, 3380–3389.

138 C. Tseng, B. Tsai, Z. Liu, T. Cheng, W. Chang and S. Lo, *Energy Convers. Manage.*, 2012, **62**, 14–21.

139 L. Li, J. Lu, H. Fang, Z. Yin, T. Wang, R. Wang, X. Fan, L. Zhao, D. Tan and Y. Wan, *IEEE Access*, 2020, **8**, 27649–27675.

140 D. Bedrov, J. Piquemal, O. Borodin, A. MacKerell Jr, B. Roux and C. Schröder, *Chem. Rev.*, 2019, **119**, 7940–7995.

141 D. Ladha, *Mater. Today Chem.*, 2019, **11**, 94–111.

142 A. Jain, Y. Shin and K. Persson, *Nat. Rev. Mater.*, 2016, **1**, 1–13.

143 L. Wang, H. Zhou, J. Hu, B. Huang, M. Sun, B. Dong, G. Zheng, Y. Huang, Y. Chen and L. Li, *Science*, 2019, **363**, 265–270.

144 S. Masoud, P. Raybaud and T. Bahers, *J. Mater. Chem. C*, 2020, **8**, 9064.

145 R. Solomon, P. Veerapandian, S. Angeline and P. Venuvanalingam, *J. Phys. Chem. A*, 2012, **116**, 4667–4677.

146 D. Whang, D. Apaydin, S. Park and N. Sariciftci, *J. Catal.*, 2018, **363**, 191–196.

147 A. Urban, D. Seo and G. Ceder, *npj Comput. Mater.*, 2016, **2**, 1–13.

148 F. Luo, A. Roy, L. Silvioli, D. A. Cullen, A. Zitolo, M. T. Sougrati, I. C. Oguz, T. Mineva, D. Teschner and S. Wagner, *Nat. Mater.*, 2020, **19**, 1215–1223.

149 A. Jain, Y. Shin and K. Persson, *Nat. Rev. Mater.*, 2016, **1**, 15004.

150 L. Fan, K. Wu, C. Tongsh, M. Zhu, X. Xie and K. Jiao, *J. Phys. Chem. Lett.*, 2019, **10**, 6409–6413.

151 L. Fan, F. Xi, X. Wang, J. Xuan and K. Jiao, *J. Electrochem. Soc.*, 2019, **166**, F511.

152 A. Muralidharan, M. Chaudhari, L. Pratt and S. Rempe, *Sci. Rep.*, 2018, **8**, 1–8.

153 S. Kim, A. Van Duin and V. B. Shenoy, *J. Power Sources*, 2011, **196**, 8590–8597.

154 R. Jinnouchi, K. Kudo, N. Kitano and Y. Morimoto, *Electrochim. Acta*, 2016, **188**, 767–776.



155 Y. Kurihara, T. Mabuchi and T. Tokumasu, *J. Electrochemical Society*, 2017, **164**, F628.

156 M. Stamatakis and D. Vlachos, *ACS Catal.*, 2012, **2**(12), 2648–2663.

157 J. Lu, H. Lei, C. Shu and C. Dai, *J. Comput. Phys.*, 2020, 109546.

158 W. Fang, Y. Tang, L. Chen, Q. Kang and W. Tao, *Int. Heat Mass Transfer*, 2018, **126**, 243–255.

159 L. Chen, Q. Kang, R. Pawar, Y. He and W. Tao, *Fuel*, 2015, **158**, 650–658.

160 S. He, B. Habte and F. Jiang, *Int. Commun. Heat Mass Transfer*, 2017, **82**, 1–8.

161 Z. Jiang, Z. Qu and L. Zhou, *Int. J. Heat Mass Transfer*, 2018, **123**, 500–513.

162 R. Groot and P. Warren, *J. Chem. Phys.*, 1997, **107**, 4423–4435.

163 Y. Ma, Y. Wang, X. Deng, G. Zhou, S. Khalid, X. Sun, W. Sun, Q. Zhou and G. Lu, *RSC Adv.*, 2017, **7**, 39676–39684.

164 Z. Niu, J. Wu, Y. Wang and K. Jiao, *J. Electrochem. Soc.*, 2018, **165**, F986.

165 L. Chen, *Annu. Rev. Mater. Res.*, 2002, **32**, 113–140.

166 Q. Li, L. Liang, K. Gerdes and L. Chen, *Appl. Phys. Lett.*, 2012, **101**, 033909.

167 C. Giménez, B. Finke, C. Schilde, L. Froböse and A. Kwade, *Powder Technol.*, 2019, **349**, 1–11.

168 R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark and F. Nerini, *Nat. Commun.*, 2020, **11**, 1–10.

169 P. De Luna, J. Wei, Y. Bengio, A. Aspuru-Guzik and E. Sargent, *Nature*, 2017, **552**, 23–27.

170 Z. Ghahramani, *Nature*, 2015, **521**, 452–459.

171 J. Schmidt, M. Marques, S. Botti and M. L. Marques, *npj Comput. Mater.*, 2019, **5**, 1–36.

172 K. Butler, D. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.

173 G. Gu, J. Noh, I. Kim and Y. Jung, *J. Mater. Chem. A*, 2019, **7**, 17096–17117.

174 D. Zhao, S. Deng, L. Zhao, W. Xu, W. Wang, X. Nie and M. Chen, *Energy AI*, 2020, 100011.

175 Y. Wang, B. Seo, B. Wang, N. Zamel, K. Jiao and X. C. Adroher, *Energy AI*, 2020, 100014.

176 B. Wang, G. Zhang, H. Wang, J. Xuan and K. Jiao, *Energy AI*, 2020, 100004.

177 Y. Wang, C. Arns, S. Rahman and J. Arns, *Math. Geosci.*, 2018, **50**, 781–799.

178 Y. Niu, P. Mostaghimi, M. Shabaninejad, P. Swietojanski and R. Armstrong, *Water Resour. Res.*, 2020, **56**, e2019WR026597.

179 H. Mostafa and X. Wang, arXiv Prepr. arXiv1902.05967.

180 Y. He, P. Liu, Z. Wang, Z. Hu and Y. Yang, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4340–4349.

181 J. Wu, X. Yin and H. Xiao, *Sci. Bull.*, 2018, **63**, 1215–1222.

182 A. Ceken, H. Dai, Y. Yabansu, S. Kalidindi and L. Song, *Acta Mater.*, 2018, **146**, 76–84.

183 R. Shams, M. Masihi, R. Boozarjomehry and M. Blunt, *J. Pet. Sci. Eng.*, 2020, **186**, 106794.

184 J. Feng, X. He, Q. Teng, C. Ren, H. Chen and Y. Li, *Phys. Rev. E*, 2019, **100**, 33308.

185 H. Chen, X. He, Q. Teng, R. E. Sheriff, J. Feng and S. Xiong, *Phys. Rev. E*, 2020, **101**, 23305.

186 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, *Advances in neural information processing systems*, 2014, pp. 2672–2680.

187 L. Mosser, O. Dubrule and M. Blunt, *Phys. Rev. E*, 2017, **96**, 43309.

188 L. Mosser, O. Dubrule and M. Blunt, arXiv Prepr. arXiv1802.05622.

189 A. Gayon-Lombardo, L. Mosser, N. Brandon and S. Cooper, *npj Comput. Mater.*, 2020, **6**, 1–11.

190 A. Valsecchi, S. Damas, C. Tubilleja and J. Arechalde, *Neurocomputing*, 2020, **399**(25), 227–236.

191 D. Jha, K. Choudhary, F. Tavazza, W. Liao, A. Choudhary, C. Campbell and A. Agrawal, *Nat. Commun.*, 2019, **10**, 1–12.

192 K. Min, B. Choi, K. Park and E. Cho, *Sci. Rep.*, 2018, **8**, 1–7.

193 M. Shandiz and R. Gauvin, *Comput. Mater. Sci.*, 2016, **117**, 270–278.

194 A. Jain, G. Hautier, C. Moore, S. Ong, C. Fischer, T. Mueller, K. Persson and G. Ceder, *Comput. Mater. Sci.*, 2011, **50**, 2295–2310.

195 G. Hautier, C. Fischer, V. Ehrlacher, A. Jain and G. Ceder, *Inorg. Chem.*, 2011, **50**, 656–663.

196 A. Jain, S. Ong, G. Hautier, W. Chen, W. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner and G. Ceder, *APL Mater.*, 2013, **1**, 11002.

197 O. Sudakov, E. Burnaev and D. Koroteev, *Comput. Geosci.*, 2019, **127**, 91–98.

198 N. Alqahtani, F. Alzubaidi, R. Armstrong, P. Swietojanski and P. Mostaghimi, *J. Pet. Sci. Eng.*, 2020, **184**, 106514.

199 Q. Rong, H. Wei, X. Huang and H. Bao, *Compos. Sci. Technol.*, 2019, **184**, 107861.

200 H. Wu, W. Fang, Q. Kang, W. Tao and R. Qiao, *Sci. Rep.*, 2019, **9**, 1–12.

201 Y. Yamamoto, M. Hattori, J. Ohyama, A. Satsuma, N. Tanaka and S. Muto, *Microscopy*, 2018, **67**, 321–330.

202 O. Furat, D. Finegan, D. Diercks, F. Viretta, K. Smith and V. Schmidt, *J. Power Sources*, 2021, **483**, 229148.

203 O. Badmos, A. Kopp, T. Bernthaler and G. Schneider, *J. Intell. Manuf.*, 2020, **31**, 885–897.

204 J. Kirman, A. Johnston, D. Kuntz, M. Askerka, Y. Gao, P. Todorovic, D. Ma, G. Prive and E. Sargent, *Matter*, 2020, **2**, 938–947.

205 T. Xie and J. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.

206 M. Karamad, R. Magar, Y. Shi, S. Siahrostami, I. Gates and A. Farimani, *Phys. Rev. Mater.*, 2020, **4**, 093801.

207 N. Jackson, A. Bowen and J. Pablo, *Macromolecules*, 2020, **53**, 482–490.

208 C. Lu, Q. Liu, Q. Sun, C. Hsieh, S. Zhang, L. Shi and C. Lee, *J. Phys. Chem. C*, 2020, **124**, 7048–7060.

209 C. Molnar, *Interpretable machine learning*, Lulu.com, 2020.

210 Q. Zhang, Y. Yang, H. Ma and Y. Wu, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6261–6270.



211 Y. Zhang and Q. Yang, *Natl. Sci. Rev.*, 2018, **5**, 30–43.

212 Y. Wang, M. Zhang, A. Lin, A. Iyer, A. Prasad, X. Li, Y. Zhang, L. Schadler, W. Chen and L. Brinson, *Mol. Syst. Des. Eng.*, 2020, **5**, 962–975.

213 B. Wu, W. Widanage, S. Yang and X. Liu, *Energy AI*, 2020, **1**, 100016.

214 Y. Wang and C. Wang, *J. Electrochem. Soc.*, 2006, **153**, A1193.

215 C. Wang, *Chem. Rev.*, 2004, **104**, 4727–4766.

216 Z. Pan, L. An, T. Zhao and Z. Tang, *Prog. Energy Combust. Sci.*, 2018, **66**, 141–175.

217 H. Xu, B. Chen, P. Tan, Y. Zhang, Q. He, Z. Wu and M. Ni, *J. Power Sources*, 2019, **440**, 227102.

218 Q. He, J. Yu, H. Xu, D. Zhao, T. Zhao and M. Ni, *Int. J. Hydrogen Energy*, 2020, **45**(22), 12459–12475.

219 H. Zhang, H. Wang and J. Xuan, *J. Power Sources*, 2020, **462**, 228113.

220 H. Zhang, H. Wang, X. Luo, D. Y. C. Leung, Q. Pang, H. Xu, L. Zhang and J. Xuan, *Chem. Eng. J.*, 2019, **356**, 673–679.

221 B. Wang, B. Xie, J. Xuan and K. Jiao, *Energy Convers. Manage.*, 2020, **205**, 112460.

222 H. Xu, J. Ma, P. Tan, B. Chen, Z. Wu, Y. Zhang, H. Wang, J. Xuan and M. Ni, *Energy AI*, 2020, 100003.

223 J. Howard, E. Tennyson, B. A. Neves and M. Leite, *Joule*, 2019, **3**, 325–337.

224 K. Severson, P. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. Chen, M. Aykol, P. Herring, D. Fragedakis, M. Bazant, S. Harris, W. Chueh and R. Braatz, *Nat. Energy*, 2019, **4**, 383–391.

225 M. Ng, J. Zhao, Q. Yan, G. Conduit and Z. Seh, *Nat. Mach. Intell.*, 2020, 1–10.

226 Y. Bengio, P. Simard and P. Frasconi, *IEEE Trans. Neural Netw.*, 1994, **5**, 157–166.

227 R. Ma, T. Yang, E. Breaz, Z. Li, P. Briois and F. Gao, *Appl. Energy*, 2018, **231**, 102–115.

228 R. Xie, R. Ma, S. Pu, L. Xu, D. Zhao and Y. Huangfu, *Energy AI*, 2020, **2**, 100017.

229 K. Jaseena and B. Kovo, *J. Renewable Sustainable Energy*, 2020, **12**, 23302.

230 Z. Deng, Y. Chen, Y. Liu and K. Kim, *Phys. Fluids*, 2019, **31**, 75108.

231 A. Khalid, A. Sundararajan, I. Acharya and A. Sarwat, 2019 *IEEE Transportation Electrification Conference and Expo (ITEC)*, IEEE, 2019, pp. 1–6.

232 Y. Jung, J. Jung, B. Kim and S. Han, *J. Clean. Prod.*, 2020, **250**, 119476.

233 R. Wang, K. Kashinath, M. Mustafa, A. Albert and R. Yu, *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 1457–1466.

234 M. Raissi, P. Perdikaris and G. Karniadakis, *J. Comput. Phys.*, 2019, **378**, 686–707.

235 J. A. Haber, Y. Cai, S. Jung, C. Xiang, S. Mitrovic, J. Jin, A. Bell and J. Gregoire, *Energy Environ. Sci.*, 2014, **7**, 682–688.

236 K. Troshin and J. Hartwig, *Science*, 2017, **357**, 175–181.

237 D. Padula, J. Simpson and A. Troisi, *Mater. Horiz.*, 2019, **6**, 343–349.

238 H. Sahu, F. Yang, X. Ye, J. Ma, W. Fang and H. Ma, *J. Mater. Chem. A*, 2019, **7**, 17480–17488.

239 H. Feng, K. Wu and Z. Deng, *Cell Rep. Phys. Sci.*, 2020, **1**(9), 100179.

240 D. Davies, K. Butler and A. Walsh, *Chem. Matter*, 2019, **31**(18), 7221–7230.

241 X. Ma, J. Lewis, Q. Yan and G. Su, *J. Phys. Chem. Lett.*, 2019, **10**(21), 6734–6740.

242 S. Sun, N. Hartono, Z. Ren and F. Oviedo, *et al.*, *Joule*, 2019, **3**(6), 1437–1451.

243 S. Back, J. Yoon, N. Tian, W. Zhong, K. Tran and Z. Ulissi, *J. Phys. Chem. Lett.*, 2019, **10**(15), 4401–4408.

244 B. Burger, P. Maffettone, V. Gusev, C. Aitchison, Y. Bai, X. Wang, X. Li, B. Alston, B. Li and R. Clowes, *Nature*, 2020, **583**, 237–241.

245 E. Gu, X. Tang, S. Langner, P. Duchstein, Y. Zhao, I. Levchuk, V. Kalancha, T. Stubhan, J. Hauch and H. Egelhaaf, *Joule*, 2020, (4), 1806–1822.

246 P. Attia, A. Grover, N. Jin, K. Severson, T. Markov, Y. Liao, M. Chen, B. Cheong, N. Perkins and Z. Yang, *Nature*, 2020, **578**, 397–402.

247 M. Duquesnoy, T. Lombardo, M. Chouchane, E. Primo and A. Franco, *ChemRxiv*. Preprint, DOI: 10.26434/chemrxiv.12473501.v1.

248 Y. Takagishi, T. Yamanaka and T. Yamaue, *Batteries*, 2019, **5**, 54.

249 T. Lombardo, J. Hoock, E. Primo, A. Ngandjong, M. Duquesnoy and A. Franco, *Batter. Supercaps*, 2020, **3**, 721–730.

250 A. Turetskyy, S. Thiede, M. Thomitzek, N. von Drachenfels, T. Pape and C. Herrmann, *Energy Technol.*, 2020, **8**, 1900136.

251 M. Popova, O. Isayev and A. Tropsha, *Sci. Adv.*, 2018, **4**, eaap7885.

252 Battery Microstructure Project, <https://made.ee.ethz.ch/research/open-source-data-and-software/battery-microstructure-project.html> (accessed January 2021).

253 Transportation Research, <https://www.nrel.gov/transportation/microstructure.html> (accessed January 2021).

254 C. Odabasi and R. Yildirim, *Energy Technol.*, 2020, **8**(12), 1901449.

255 A. Torayev, P. Magusin, C. Grey, C. Merlet and A. Franco, *J. Phys.: Mater.*, 2019, **2**, 044004.

256 J. Janet, S. Ramesh, C. Duan and H. Kulik, *ACS Cent. Sci.*, 2020, **6**(4), 513–524.

257 K. Weiss, T. Khoshgoftaar and D. Wang, *J. Big Data*, 2016, **3**, 9.

258 J. Janet and H. Kulik, *J. Phys. Chem. A*, 2017, **121**(46), 8939–8954.

259 M. A. Bessa, P. Glowacki and M. Houlder, *Adv. Mater.*, 2019, **31**, 1904845.

260 J. Granda, L. Donina, V. Dragone, D. Long and L. Cronin, *Nature*, 2018, **559**, 377–381.

