Check for updates

Introduction to "Accelerate Conference 2022"

Keith A. Brown, [ID] [a] Fedwa El Mellouhi [ID] [b] and Claudiane Ouellet-Plamondon [ID] [c]

This new themed collection represents a collaboration between the editors of *Digital Discovery* and the organisers of the 2022 Accelerate Conference. The goal of the conference was to explore the power of self-driving labs (SDLs), which combine AI, automation, and advanced computing to accelerate materials and molecular discovery. This themed collection features contributions on this topic from researchers who presented at that meeting.

Vescovi *et al.* introduce the concept of science factories, large-scale self-driving laboratories powered by simulation and AI technologies (https://doi.org/10.1039/D3DD00142C). The paper outlines methods for modular construction, grouping modules into work cells, and executing applications across these units. They review 15 experiments with robotic apparatus and five applications (one in education, two in materials, and two in biology). The modular architecture accommodates diverse instruments, AI models, and data repositories, enabling versatile experimentation. The architecture's modularity facilitates rapid adaptation and collaboration, with an emphasis on performance optimization and

scalability. Future developments aim to upscale simulation capacities for modeling complex operational dynamics and validating large-scale deployment feasibility. The flexibility and scalability of science factories offer a pathway to addressing diverse scientific challenges and advancing knowledge across multiple domains. Collaboration is encouraged to refine and expand its capabilities for broader scientific impact.

In contrast to centralized approaches that envision a future in which large facilities perform experiments for users, another path is to realize low-cost self-driving labs that can be replicated in many labs around the world. Key to this process is the development of reliable and open-source hardware that can perform many different types of experiments. As an example of this, Politi *et al.* report a new platform to study the synthesis of quantum dot nanoparticles (https://doi.org/10.1039/D3DD00033H). This process combines autonomous reagent dispensing using a robotic pipetting system, synthesis using ultrasonic cavitation, and optical characterization. Notably, this approach uses the same well plate for formulation, synthesis, and characterization, which enabled the team to study >600 distinct formulations. In addition to allowing them to identify optimized synthesis conditions, this process represents

a facile extension of an open-source hardware platform, namely the Jubilee system, in a manner that shows its promise for adaptation to diverse chemical tasks.

The development of high-throughput platforms for studying chemical processes requires innovations in both hardware and chemical processes that are sufficiently general to explore a wide range of materials. Lee *et al.* (https://doi.org/10.1039/D2DD00100D) demonstrate a fully automated system that enables the high-throughput exploration of PET-RAFT polymerization, which is a common and powerful approach for forming libraries of polymers with applications ranging from biology to functional materials. To do this, they combine robotic sample dispensing with a custom-multiplexed illumination system that allows each of the 96 wells in the plate array to be independently illuminated and studied using periodic optical characterization. This work shows the importance of the codesign of hardware and chemical synthesis pathways to realize versatile self-driving labs for exploring polymer formulations.

Processing of solid materials such as a polymers is subject to a distinct set of needs from liquid-phase processing, which are often less amenable to automation. In Hernandez-del-Valle *et al.*, an automated material acceleration

aBoston University, USA

bQatar Environment and Energy Research Institute, Hamad Bin Khalifa University, Qatar

cÉcole de Technologie Supérieure, Université du Québec, Canada

platform is reported that enables the development of new thermoplastic bio-based PLA materials for 3D printing (https://doi.org/10.1039/D3DD00141E). This work shows the autonomous optimization of process conditions through optimization of print quality control by adjusting the printing speed modifier, the layer height and layer width, and the extrusion flow multiplier. The robotized workflow inputs the polymer pellet, identifies 3D printing parameters, and performs mechanical property testing with the Charpy test. The design of experiments based on Latin hypercube sampling identified specimens of sufficient quality in 10 experiments, and these were further improved with Bayesian optimization.

While novel hardware is important, coordination is needed to make sure the community moves forward productively. Maffettone et al. urge the community to prioritize actions that ensure the adaptability and interoperability of hardware within self-driving laboratories in their contribution (https://doi.org/10.1039/D3DD00143A). They highlight open-source hardware communication and standardized interfaces for both data and physical sample exchange. Although their contribution advocates for the publication of modular hardware components, it also notes that anthropomorphic robots can allow the integration of bespoke equipment.

Pelkie and Pozzo discuss their perspective on the community's need for integrated materials data management (https://doi.org/10.1039/D3DD00022B). Effective data flow management is required throughout the data life cycle, which in their framework progresses from experimental data collection systems through group data management systems to dissemination in data-sharing community ecosystems. They emphasise that open-source tools create new workflows, facilitate community engagement, and connect the individual labs to a network of databases. Broad community buy-in, including by technical, scientific and governmental organizations, is needed to realize these goals.

Snapp and Brown outline a novel approach to training and operating self-driving laboratories, drawing from experience of the Bayesian experimental autonomous researcher (BEAR) program (https://doi.org/10.1039/D3DD00150D). Over a span of two years, the BEAR initiative conducted over 25 000 experiments, employing a highly dynamic process that continuously adjusted settings based on real-time feedback from the SDL's progress. The manuscript delves into the key insights derived from this campaign, focusing on two critical aspects: the decision-making process for experimenters and the essential information required to inform these decisions. Six pivotal settings and four key plots are identified as central to optimizing the performance of SDLs throughout a campaign. These settings and monitoring metrics transcend specific research domains, offering valuable lessons applicable across various scientific disciplines. Leveraging the widely accepted framework of Bayesian optimization, they offer a structured approach to implementing these insights. While acknowledging that every SDL operation may present unique challenges, the principles and processes elucidated here aim to enhance experimenters' intuition and efficacy in managing SDLs. By offering a standardized language and framework, the paper seeks to democratize the use of SDLs, particularly within the materials science domain.

One hallmark of materials data is that it is high-dimensional. As a result, it can be extremely challenging to separate materials into distinct categories. In order to address this difficulty, Bonakala et al. explore novel interactive approaches for evaluating and identifying classes of materials (https://doi.org/10.1039/D3DD00179B). They show that conventional unsupervised classification techniques are not readily able to partition metal–organic frameworks (MOFs) into meaningful classes, at least in two dimensions. Instead, they share a novel divide-and-conquer approach that combines Gaussian mixture models and eigenvalue decomposition discriminant analysis and proposes many possible two-dimensional plots. At this point, the user decides to either merge or split each set of classes using a series of visual decisions. Finally, the resulting material classes are represented as a graph to complete the classification process. Crucially, this work merges key benefits of data-driven work in terms of rapidly interpreting high-dimensional space with human intuition interjected at a key stage. As such, this process provides a thought-provoking example of human–machine partnership in the era of data-driven research.

Schopmans et al. developed an algorithm to generate synthetic crystals from ICSD powder X-ray diffractograms to assist in model training (https://doi.org/10.1039/D3DD00071K). Traditional methods and the current ICSD database face challenges due to limited data availability and class imbalances. They proposed a distributed computing architecture to build an infinite stream of synthetically generated and simulated diffractograms. They used deep convolutional neural networks, trained on millions of unique synthetic diffractograms to classify diffractograms' space groups. This model trained on synthetic data had a better group classification accuracy than one trained on ICSD structures. This indicates the potential of using synthetically generated crystals to extract structural information, enabling the application of large machine learning models in powder X-ray diffraction.

Joress et al. propose an automated cementitious-material formulation and testing platform motivated by the critical need to repair infrastructure in the USA and worldwide (https://doi.org/10.1039/D3DD00211J). The machine learning (ML) agent can take three levels of decision: the solid mixture; the hydration ratio and other liquid admixtures; and the tests to perform. The automated material characterization must occur on uncured paste, curing cement, and cured products. The ML agent will likely comprise Gaussian process models to allow identifying unexplored regions of the parameter space. More advanced models can then boost the prediction performance. The envisioned robotic platform will allow the validation and

optimization of the formulation of new specialized repair materials.

This collection draws together the meeting's themes of the realization of new SDLs; fundamental studies of the operation of SDLs; and sustainable, resilient, low-carbon, materials and chemical discoveries made using SDLs, AI, and machine learning. We thank the contributors for their work, and hope that you enjoy reading this collection. These topics will be further expanded on in further Accelerate conferences.

Acknowledgements