An object-oriented framework to enable workflow evolution across materials acceleration platforms

Leong, Chang Jie; Low, Andre Kai Yuan; Recatala-Gomez, Jose; Velasco, Pablo Quijano; Vissol-Gaudin, Eleonore; Tan, Jin Da; Ramalingam, Balamurugan; Made, Riko I; Pethe, Shreyas Dinesh; Sebastian, Saumya; Lim, Yee-Fun; Khoo, Jonathan Zi Hui; Bai, Yang; Cheng, Jayce Jian Wei; Hippalgaonkar, Kedar

2022

Leong, C. J., Low, A. K. Y., Recatala-Gomez, J., Velasco, P. Q., Vissol-Gaudin, E., Tan, J. D., Ramalingam, B., Made, R. I., Pethe, S. D., Sebastian, S., Lim, Y., Khoo, J. Z. H., Bai, Y., Cheng, J. J. W. & Hippalgaonkar, K. (2022). An object-oriented framework to enable workflow evolution across materials acceleration platforms. Matter, 5(10), 3124-3134. https://dx.doi.org/10.1016/j.matt.2022.08.017

https://hdl.handle.net/10356/164443

https://doi.org/10.1016/j.matt.2022.08.017

An Object-Oriented Framework to Enable Workflow Evolution across Materials Acceleration Platforms

Chang Jie Leong[#1], Kai Yuan Andre Low[#2], Jose Recatala-Gomez[#2], Pablo Quijano Velasco[#1], Eleonore Vissol-Gaudin[2], Jin Da Tan[1,4], Balamurugan Ramalingam[1,3], Riko I Made[1], Shreyas Dinesh Pethe[2], Saumya Sebastian[2], Yee-Fun Lim[1,3], Zi Hui Jonathan Khoo[1], Yang Bai[1], Jayce Jian Wei Cheng[1], Kedar Hippalgaonkar[1,2*]

[1]Institute of Materials Research and Engineering, Agency for Science Technology and Research, 2 Fusionopolis Way, #08-03 Innovis, 138634, Singapore.

[2]School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Block N4.1, 639798, Singapore.

[3]Institute of Sustainability for Chemicals, Energy and Environment, Agency for Science Technology and Research, 1 Pesek Rd, 627833, Singapore

[4]National University of Singapore Graduate School – Integrative Sciences and Engineering Programme, 21 Lower Kent Ridge Road, 119077, Singapore

[#] Denotes equal contribution.

[*]Correspondence: kedar@ntu.edu.sg

Summary

Progress in data-driven self-driving laboratories for solving materials grand challenges has accelerated with the advent of machine learning, robotics and automation but usually designed with specific materials and processes in mind. To develop the next generation of Materials Acceleration Platforms (MAPs), we propose a unified framework to enable collaboration between MAPs, leveraging on object-oriented programming principles using which research groups around the world would be able to effectively evolve experimental workflows. We demonstrate the framework via three experimental case studies from disparate fields to illustrate the evolution of, and seamless integration between workflows, promoting efficient resource utilisation and collaboration. Moving forward, we project our framework on three other research areas that would benefit from such an evolving workflow. Through the wide adoption of our framework, we envision a collaborative, connected, global community of MAPs working together to solve scientific grand challenges.

Introduction

Novel materials and devices are often the impetus behind scientific progress in fulfilling societal and technological needs. Historically, materials development has been slow, with most technologies taking at least 20 years from discovery to commercialization [1]. Such a time frame is not adequate to alleviate the problems posed by climate change and a growing energy economy. To meet these needs, accelerated materials innovation has been achieved through synergistic combination of advancements in automation and high-throughput (HT) capabilities, as well as integration of machine learning and artificial intelligence for active learning and closed-loop experimentation [2–9]. Such infrastructure has culminated in the development of self-driving laboratories or Materials Acceleration Platforms (MAPs), which hold the key to efficiently navigating the vast chemical space of functional and structural materials. This approach has already proved to be successful for the discovery of materials relevant to tackling climate change such as perovskite solar cells, solar fuel photoanodes, organic polymer photocatalysts among many others [10–13].

With countless possible materials, compounded by the multitude of different possible manufacturing and characterization processes, the full potential of these MAPs is yet to be unlocked. Currently, individual MAPs tend to be developed with self-contained experimental workflows in mind for a targeted material or application. The development of a MAP tends to be a time-intensive process where the workflow evolves through domain knowledge and engineering constraints. This leads to limitations within the community where MAPs, and their associated research groups, tend to operate in isolation. A new framework is required to allow for interfacing among different MAPs with evolving workflows and experimental capabilities, characterization techniques, and analytics, so that cross-pollinated data-driven experiments can be conceived and executed. This can enable the flexible execution for collaborators to share domain knowledge, data, tools, and processes, and integrate these innovations within a global community.

In response to the call-to-action by Seifrid et al. [14] for ideas to develop the next-generation MAPs, we present in this white paper a framework inspired by object-oriented programming principles to enable such evolution of processes and materials discovery via collaborative experimental workflows. In the next section, we elaborate on our framework and illustrate a typical materials development workflow that can be executed with collaborative MAPs.

Proposal

In an object-oriented paradigm[15], *classes* are user-defined blueprints from which *objects* are instantiated. Methods are functions defined in classes, which is an abstraction of an action, to maintain reusability or encapsulate functionality within the class or object. *Attributes* describe information or features that a class can contain, with specifically assigned values in objects. Borrowing from this paradigm, we envision a workflow planning framework for MAP development resembling software code with well-defined input and return variable types.

In our framework, as depicted in Figure 1, a workflow comprises a sequence of *states* linked by a series of *tasks* from initial to final state, which we define as the base classes. These states can be physical samples or data, that we keep records of, or collect data from. Examples of states are chemical precursors, material properties, and trained models. An *article* is an instance of the state class. One could start by identifying the final output state and work backwards to identify the intermediate states and inputs to acquire. These states serve as checkpoints and the path between checkpoints can change, through intelligent planning, which we discuss further below.

Tasks can be broken down further into *actions* by specifying the use of a particular *tool*. We define four categories of tasks: material synthesis (chemical reactions/formulations), material processing (physical transformations), characterization (measurements), and analysis (cognitive functions). Thereafter, a *technique* is a subclass of a task (e.g., material processing ⊃ thin film deposition ⊃ spin coating). There can be many permutations of techniques and tools for a given task, where some techniques naturally limit themselves to particular tools. Techniques can then be selected based on constraints placed by the input and output state attributes, and inherit similar attributes from the same parent task. A tool that can complete a task using a specific technique is an instance of that technique subclass, and these objects contain methods that we refer to as actions. From these tools, we can collect associated metadata.

We make use of two key principles of object-oriented thinking: abstraction and inheritance. Through abstraction, we distil the states and tasks into their simplest tool-agnostic forms. For states, abstraction outlines the scope of objects we might encounter in the workflow, surfacing only the relevant attributes and properties for defining tasks. These state attributes narrow down the possible tasks that are reconcilable with a given input/output pair. For tasks, abstraction creates the flexibility to continually evolve tools used to accomplish the tasks, ensuring compatibility with the rest of the workflow since the input and output states match.

With inheritance, we establish a clear hierarchy of classes and lay out the common attributes and methods at each level. For states, this is useful for describing the range of objects handled in that node in enough detail, without loss of generality. A state can then be defined with increasing specificity by

creating subclasses, which inherit the attributes of the base class (e.g., liquid ⊃ solvent ⊃ reagent). For tasks, having this hierarchy brings out the common control parameters and attributes across a suite of subclasses (i.e., techniques) or objects (i.e., tools). This can help us describe specific limitations of the technique/tool. Table 1 provides a glossary of the terms/concepts we defined beforehand.

To perform intelligent planning of experimental workflow that optimises task metrics with the same desired intermediate states and final output, we need sufficient metadata that describes the various 'routes' (i.e., from different MAPs of differing capabilities) available to prepare an identical sample. This can be described by the job-shop scheduling problem [17]: there is a specific order of tasks to be completed (e.g., synthesis to characterization to modelling), where constraints can be set such that certain tasks are restricted to certain machines. The goal is to then optimise for the aforementioned task metrics as described in Figure 1, such that we derive an evolved workflow [18]. This can be computationally solved via optimization and scheduling algorithms [19,20]

In line with the FAIR data principles [21], we envisage two distinct sets of data being generated during the evolution of a workflow: (1) the materials and tool-specific process data that relates to the inputs and outputs of an experiment and used for closed-loop optimisation, and (2) the attributes / methods of the various states and tasks that constitute a workflow using our object-oriented framework. Current practices of storing structured data in relational databases remains best-suited to link materials and tool-specific process data together, since it encapsulates all the relevant information for closed-loop machine learning applications. However, using a relational database to store the relevant information of the states and tasks used in our object-oriented framework would result in a sparse table as there is a huge diversity in the attributes and methods that define each instance of a class. Non-tabular data files such as JSON files are more suitable to provide flexibility to store the attributes / methods of different states and tasks. Implementing object-relational or NoSQL databases would provide a solution for this mismatch. PostgreSQL is an object-relational database management system that supports the SQL capabilities relevant to store data type (1) and offers flexible non-relational storage of information as JSON files for data type (2).

Finally, the application of this framework can highlight tooling gaps between states and the opportunity for collaboration between teams. With predefined inputs and outputs serving as a "handshake" or an interface between different MAPs or tools, different teams can work on different tasks or work on improving the same technique concurrently, knowing that they can be coupled.

Case studies

Here, we present three case studies to illustrate how our object-oriented framework can be used to enable the evolution of workflows.

CS1: Enabling accurate MWD prediction in inaccessible chemical space

In this case study, we demonstrate our evolved workflow to create a model that predicts Molecular Weight Distribution (MWD) of polystyrene for a given set of input reaction parameters such as monomer and initiator concentrations [22]. Figure 2 shows our workflow mapping from the initial state represented by the precursors class to the output state which is a prediction of polystyrene MWD.

Our initial aim was to use a computer-controlled segmented-flow reactor to carry out multiple experimental conditions at once in a HT fashion. As we proceeded to explore the entire parameter space on the flow reactor, we observed precipitation of AIBN in the syringes for reactions that had both high AIBN to styrene ratios and high styrene concentrations. Rather than be bound by engineering limitations of the flow reactor, we evolved the experimental workflow by incorporating a batch reactor that met the requirements of the inaccessible chemical space. Thus, we carried out a subset of reaction conditions using a batch technique that would complement the MWD data of HT flow technique. The second task shown in Figure 2 describes this evolution in our workflow. Nonetheless, the abstracted workflow still allows for future improvements of our material synthesis tools into improved versions that completely rely on automation and can explore the full parameter space.

For the analysis of our MWD data our workflow had to evolve to be able to utilise the data obtained from two different material synthesis techniques. Thus, we used transfer learning from the HT flow reactor to batch experiments enhancing the predictive ability of the Deep Neural Network (DNN) algorithm. Therefore, we pre-trained a DNN using data from the polymers generated from the flow technique, freezing all the weights of the DNN except for the last layer and allowing it to be fine-tuned using data from the polymers generated from the batch technique. The resultant predictions were highly accurate. Through this work, we demonstrated the advantages of having parallel workflows (flow and batch), in which other research groups can also integrate into their MAPs, or utilise existing setups like ours.

CS2: Tuning quantum dot (QD) surface ligands for high electrical performance

In the next case study, we investigated the effect of ligand mixtures on the electrical conductivity of QD thin films. We first predefined the sequence of input and output states in our workflow, as well as the linking tasks (Figure 3). Given that we could deposit patterned device contacts in batches of 20, our bottlenecks were in fabricating thin films and characterising devices.

The QD deposition task involves 5 different reagents across 7 separate spin steps, including ligand exchange, surface treatment, and washing steps. To get a consistent 100 nm thin film from a liquid starting material, we selected a spin-coating technique. When we initially did it manually (with a Laurell

spin-coater), much of each 1-minute spin step was spent waiting. Hence, we built a multi-spinner tool to concurrently process 4 substrates, increasing throughput with parallelisation. To eliminate errors from changing pipette tips and increase repeatability, we coupled it with an automated multi-channel dispensing setup. These added up to give our desired throughput and film quality.

For electrical characterisation, we used a 2-point probe setup connected to two Keithley 2450 machines, one each for source-drain and gate voltages, and manually made contact with the sample devices. Each substrate had several devices of differing channel lengths, making probe alignment after each voltage sweep a time-consuming process. To tackle this, we built a 3-axis cartesian robot to place the probes on the device contacts, enabled by a machine-vision system that identifies the device positions and automatically moves the probes to the next device after each measurement. In both tasks, the framework pinpoints potential areas for process evolution and the well-defined input/output states ensure compatibility of the improved tools with the rest of the workflow. Furthermore, our work identified and corrected an engineering gap within QD thin films, which could be shared with other thin film MAPs.

CS3: Solid-state synthesis

This case study describes the application of an object-oriented framework to solid-state synthesis (Figure 4), which involves long range diffusion and rearrangement of atomic species [23,24]. Due to the sluggish solid diffusion processes, high temperature or pressure are held for a long time to overcome the kinetic barriers for diffusion [25]. Solid-state synthesis can be abstracted into two tasks: i) reagent mixing and ii) solid-state reaction, with the latter usually being the rate-limiting task.

The first task (mixing) acts on the initial state (elemental precursors) using a ball milling tool, changing the subclass. Subsequently, state 2 was heated up (action 1) to target temperature using a manual tool for phase formation. The process evolved targeting a different subclass of state 1: precursor homogenization at the nanoscale [26]. This was achieved by switching the tool used in action 1, from ball milling to spark plasma discharge. Although the diffusion time is indeed reduced, we find that the bottleneck is not task 1 but task 2, specifically the slow ramping and cooling rates.

The tool used in task 2 was switched from furnace to rapid joule heating, resulting in state 3, an inorganic compound. The next task is to elucidate the crystal structure, carried out using the x-ray diffraction (XRD) technique. Further process evolution can be achieved by switching the tool to a synchrotron facility, increasing throughput and fidelity, for instance through collaboration with MAPs such as the Stanford Linear Accelerator Center. Analysis, for example, combines the Fourier-transformed crystal properties (FTCP) algorithm for encoding crystal structure and transfer learning to predicted structure in state 5. The predictions of the model are then fed back into the workflow (new state 1) and another iteration of the workflow is carried out.

Future works

Our case studies demonstrate how evolving workflows might become a necessity given the increasing complexity of experiments and the inherent limitations of the tools used within a MAP. Instead of restricting experimentalists to only a small subset of the parameter space, we should strive to be flexible and adapt to changing requirements of a particular workflow. Below, we show three examples that could benefit from using our proposed framework to map out the possible tasks that could be used to maximise the experimental space to arrive at a specific final state.

FW1: Metal oxide synthesis for photo/electrocatalytic materials

Sol-gel synthesis in mild (low temperature and pressure) conditions [27] ushers in the possibility of HT experiments for metal oxide film discovery. Our group recently employed a MAP to investigate Cu-Sb-S oxide film materials, integrating a self-driving HT robot to automate tasks such as precursor mixing, spin-coating, and annealing. All this was implemented in closed-loop experimentation via Bayesian Optimization to drive active learning. It only took several weeks to converge on a composition-process solution with ~100 multi-labelled experiments. However, this implementation had its limitations including the inability to handle high-viscosity liquids, spin coating one sample at a time, and offline characterization of the samples.

An evolution of the mixing task could involve integrating an overhead gravimetric dispenser and robots that can handle highly viscous liquids. Additionally, we can simplify the characterization task by using a scanning droplet cell (SDC), which directly measures photo/electrocatalytic performance. This replaces absorption, XPS and film uniformity measurements that served as proxies for photocurrent. Automated SDCs [28] and multi-channel cells can further accelerate catalytic measurement. With these tool upgrades, we seek to rapidly synthesise, characterise, and measure thousands of multi-metal oxides with the newly evolved workflow. Massive data libraries could subsequently be quickly generated to explore the genome of multi-metal oxides for catalysis.

FW2: Metal Halide Perovskite Opto-electronics

Considerable research has been conducted on using MAPs for metal halide perovskite research, such as for investigating anti-solvents [29–31], solution synthesis of perovskites [32–35], thin film processing of perovskites [36,37], and device fabrication. Perovskites are amenable to solution processing, and the vast compositional space for exploration makes them an ideal candidate for HT studies. In addition, HT characterization can readily be done with absorbance and photoluminescence measurement instruments.

We can consider various liquid handling systems that automate the formulation task to study different perovskite compositions. Different film fabrication techniques like spin-coating, drop-casting, blade coating, and slot-die coating can be explored to meet our throughput and film quality requirements. HT opto-electrical characterisation tools such as a robot-arm-mounted absorbance and PL detector, or hyperspectral imaging could be incorporated. Using well-defined states and tasks, we can create reconfigurable MAPs that facilitate inter-group collaboration and accelerate the perovskite development. This is a noted limitation of existing commercial systems, which while sophisticated and very high in throughput, are narrow in scope with poor modularity. We propose collaboration with MAPs such as AMANDA [7] that allow for addition of further functionalities and workflow evolution, by integrating different modules and software capabilities.

FW3: Unconventional computing device development and discovery

Our proposed framework can be extended to fields of research relying on the development and exploitation of novel material systems, such as Unconventional Computing (UC) [38,39] and its subfields, including Neuromorphic Computing and Evolution in Materio (EiM) [40,41]. Despite various efforts to produce common collaborative frameworks [42,43], advances are hindered by lack of comparability in results obtained by different groups. This is mainly due to the wide range of possible material concentration-process-architecture-algorithm combinations and the lack of common characterisation benchmarks. Our framework could address these issues by providing a unifying means of developing experimental UC workflows.

EiM for example, which aims to explore and exploit material systems to solve non-trivial computational problems, could follow a supervised [44], or unsupervised [43,45] workflow. A supervised experimental workflow would have a material processor as its target. Materials, states, and interfaces would be articles evolved through combinatorial search. Material synthesis and processing such as mixing and deposition could be evolved by moving to automated and HT techniques. Finally, characterisation would be split between using standardised techniques and tools (Source-meter, SEM, etc) and using techniques and tools specific to the target (Optimisation Algorithms, Mecobo, etc). This would result in the development of databases containing primary characteristics (electrical, optical, etc) and secondary characteristics (memory, efficiency, etc) of a wide range of materials and composites, and crucially, the metadata associated with processes and tools used to produce them. When shared and exploited by different research groups, this would enable the analysis, optimisation, and production of new EiM devices.

Limitations

While having MAPs and inter-group collaboration bring about many benefits, we recognise certain shortcomings that our framework could overcome. Firstly, the development and maintenance cost of MAPs is significant, it is therefore necessary to establish clear motivations and tool gaps to justify their implementation into our framework for global collaboration. Secondly, process drift and equipment breakdowns will happen with MAPs, as with any equipment. There needs to be clearly defined attributes for states and tasks within each MAP such that they can act as auxiliaries for each other to minimise research downtime. Finally, there are logistical concerns with travel time and sample stability when attempting collaboration across multiple MAPs. This has severe implications for certain materials, such as time-dependent ageing/oxidising, or simply due to the fragility of certain devices such as those in thin-films which makes transportation difficult. Defining stability and measurement standards as key metrics for the 'transportation' task within our framework enables proper sample transfer to occur, thereby allowing for seamless collaboration.

Conclusion

To fully meet their potential, next-generation MAPs need to be integrated into an interconnected global network that facilitates the exchange of digital (data and software) and physical (hardware components, samples, measurement techniques) resources. In this white paper we have proposed to take advantage of object-oriented programming principles to build such a framework, with self-consistent states and tasks to construct well-defined workflows. By using the properties of abstraction and inheritance, we describe how flexible and collaborative workflows may be constructed that use a variety of techniques to arrive at the same final state. Finally, through three case studies and three areas of future work, we demonstrate the benefit of such an approach for the planning of evolving workflows for materials discovery. We believe that our framework provides a blueprint for a common language to share experiments and equipment that provides the flexibility for global collaboration, which will enable the transition away from isolation to a hyper-connected research landscape for the next generation of MAPs.

Author contributions

All authors contributed to the conceptualization and writing of this manuscript.

Acknowledgements

We acknowledge funding from Accelerated Materials Development for Manufacturing Program A1898b0043 at A*STAR via the AME Programmatic Fund by the Agency for Science, Technology

and Research. K.H. also acknowledges funding from the NRF Fellowship NRF-NRFF13-2021-0011.

Declaration of Interests

K.H. owns equity in a start-up focused on accelerating materials development through machine learning and high-throughput experimentation.

References

1. Correa-Baena, J.-P., Hippalgaonkar, K., van Duren, J., Jaffer, S., Chandrasekhar, V.R., Stevanovic, V., Wadia, C., Guha, S., and Buonassisi, T. (2018). Accelerating Materials Development via Automation, Machine Learning, and High-Performance Computing. Joule *2*. 1410–1420.

2. Lookman, T., Balachandran, P. v, Xue, D., Pilania, G., Shearman, T., Theiler, J., Gubernatis, J.E., Hogden, J., Barros, K., and BenNaim, E. (2016). A perspective on materials informatics: state-of-the-art and challenges. In Information science for materials discovery and design (Springer), pp. 3–12. 10.1007/978-3-319-23871-5_1.

3. Lookman, T., Alexander, F.J., and Rajan, K. (2016). Information science for materials discovery and design (Springer). 10.1007/978-3-319-23871-5.

4. Liu, Y., Zhao, T., Ju, W., and Shi, S. (2017). Materials discovery and design using machine learning. Journal of Materiomics *3*, 159–177. 10.1016/j.jmat.2017.08.002.

5. Coley, C.W., Thomas III, D.A., Lummiss, J.A.M., Jaworski, J.N., Breen, C.P., Schultz, V., Hart, T., Fishman, J.S., Rogers, L., and Gao, H. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. Science (1979) *365*. 10.1126/science.aax1566.

6. Burger, B., Maffettone, P.M., Gusev, V. v, Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., and Clowes, R. (2020). A mobile robotic chemist. Nature *583*, 237–241. 10.1038/s41586-020-2442-2.

7. Wagner, J., Berger, C.G., Du, X., Stubhan, T., Hauch, J.A., and Brabec, C.J. (2021). The evolution of Materials Acceleration Platforms: toward the laboratory of the future with AMANDA. Journal of Materials Science *56*, 16422–16446. 10.48550/arXiv.2104.07455.

8. MacLeod, B.P., Parlane, F.G.L., Morrissey, T.D., Häse, F., Roch, L.M., Dettelbach, K.E., Moreira, R., Yunker, L.P.E., Rooney, M.B., and Deeth, J.R. (2020). Self-driving laboratory for accelerated discovery of thin-film materials. Science Advances *6*. 10.1126/sciadv.aaz8867.

9. Lusann, Y., A, H.J., Zan, A., J, Y.S., Kevin, K., Lan, Z., H, R.M., Christopher, R., Nicholas, W., Marc, C., et al. (2021). Discovery of complex oxides via automated experiments and data science. Proceedings of the National Academy of Sciences *118*. 10.1073/pnas.2106042118.

10. Bai, Y., Wilbraham, L., Slater, B.J., Zwijnenburg, M.A., Sprick, R.S., and Cooper, A.I. (2019). Accelerated discovery of organic polymer photocatalysts for hydrogen evolution from water through the integration of experiment and theory. J Am Chem Soc *141*, 9063–9071. 10.1021/jacs.9b03591.

11. Sun, S., Hartono, N.T.P., Ren, Z.D., Oviedo, F., Buscemi, A.M., Layurova, M., Chen, D.X., Ogunfunmi, T., Thapa, J., and Ramasamy, S. (2019). Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis. Joule *3*, 1437–1451. 10.1016/j.joule.2019.05.014.

12. Yan, Q., Yu, J., Suram, S.K., Zhou, L., Shinde, A., Newhouse, P.F., Chen, W., Li, G., Persson, K.A., and Gregoire, J.M. (2017). Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment. Proceedings of the National Academy of Sciences *114*, 3040–3043. 10.1073/pnas.1619940114.

13. Stach, E., DeCost, B., Kusne, A.G., Hattrick-Simpers, J., Brown, K.A., Reyes, K.G., Schrier, J., Billinge, S., Buonassisi, T., Foster, I., et al. (2021). Autonomous experimentation systems for materials development: A community perspective. Matter *4*, 2702–2726. 10.1016/j.matt.2021.06.036.

14. Seifrid, M., Hattrick-Simpers, J., Aspuru-Guzik, A., Kalil, T., and Cranford, S. (2022). Reaching critical MASS: Crowdsourcing designs for the next generation of materials acceleration platforms. Matter. 10.1016/j.matt.2022.05.035.

15. Larman, C. (2005). Applying UML and Patterns: An Introduction to Object-oriented Analysis and Design and Iterative Development (Pearson).

16. Ng, D.Z.L., Nelson, A.Z., Ward, G., Lai, D., Doyle, P.S., and Khan, S.A. (2022). Control of Drug-Excipient Particle Attributes with Droplet Microfluidic-based Extractive Solidification Enables Improved Powder Rheology. Pharmaceutical Research *39*, 411–421. 10.1007/s11095-021-03155-0.

17. MacCarthy, B., and Liu, J. (1993). Addressing the gap in scheduling research: a review of optimization and heuristic methods in production scheduling. International Journal of Production Research - INT J PROD RES *31*, 59–79. 10.1080/00207549308956713.

18. Malakooti, B. (2014). Operations and production systems with multiple objectives (John Wiley & Sons).

19. Chen, X., Lan, Y., Benko, A., Dósa, G., and Han, X. (2011). Optimal algorithms for online scheduling with bounded rearrangement at the end. Theoretical Computer Science *412*, 6269–6278. 10.1016/j.tcs.2011.07.014.

20. Khuri, S., and Miryala, S. (1999). Genetic Algorithms for Solving Open Shop Scheduling Problems. 10.1007/3-540-48159-1_25.

21. Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data *3*, 160018. 10.1038/sdata.2016.18.

22. Tan, J. da, Ramalingam, B., Liang Wong, S., Cheng, J., Lim, Y.-F., Chellappan, V., Khan, S.A., Kumar, J., and Hippalgaonkar, K. (2022). Machine Learning Predicts Conversion and Molecular Weight Distributions in Computer Controlled Polymerization. 10.26434/chemrxiv-2022-tlz53.

23. Jansen, M. (2002). A Concept for Synthesis Planning in Solid-State Chemistry. Angewandte Chemie International Edition *41*, 3746–3766. 10.1002/1521-3773(20021018)41:20%3C3746::AID-ANIE3746%3E3.0.CO;2-2.

24. Shewmon, P. (2016). Diffusion in solids (Springer).

25. Schön, J.C., and Jansen, M. (1996). First Step Towards Planning of Syntheses in Solid-State Chemistry: Determination of Promising Structure Candidates by Global Optimization. Angewandte Chemie International Edition in English *35*, 1286–1304. 10.1002/anie.199612861.

26. Qi, W.H., and Wang, M.P. (2004). Size and shape dependent melting temperature of metallic nanoparticles. Materials Chemistry and Physics *88*, 280–284. 10.1016/j.matchemphys.2004.04.026.

27. Hench, L.L., and West, J.K. (1990). The sol-gel process. Chemical Reviews *90*, 33–72. 10.1021/cr00099a003.

28. Gregoire, J.M., Xiang, C., Liu, X., Marcin, M., and Jin, J. (2013). Scanning droplet cell for high throughput electrochemical and photoelectrochemical measurements. Review of Scientific Instruments *84*, 024102. 10.1063/1.4790419.

29. Gu, E., Tang, X., Langner, S., Duchstein, P., Zhao, Y., Levchuk, I., Kalancha, V., Stubhan, T., Hauch, J., Egelhaaf, H.J., et al. (2020). Robot-Based High-Throughput Screening of Antisolvents for Lead Halide Perovskites. Joule *4*, 1806–1822. 10.1016/j.joule.2020.06.013.

30. Higgins, K., Valleti, S.M., Ziatdinov, M., Kalinin, S. v, and Ahmadi, M. (2020). Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning. ACS Energy Letters *5*, 3426–3436. 10.1021/acsenergylett.0c01749.

31. Manion, J.G., Proppe, A.H., Hicks, G.E.J., Sargent, E.H., and Seferos, D.S. (2020). High-Throughput Screening of Antisolvents for the Deposition of High-Quality Perovskite Thin Films. ACS Applied Materials & Interfaces *12*, 26026–26032. 10.1021/acsami.0c06110.

32. Abdel-Latif, K., Epps, R.W., Bateni, F., Han, S., Reyes, K.G., and Abolhasani, M. (2021). Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow. Advanced Intelligent Systems *3*, 2000245. 10.1002/aisy.202000245.

33. Bateni, F., Epps, R.W., Antami, K., Dargis, R., Bennett, J.A., Reyes, K.G., and Abolhasani, M. (2022). Autonomous Nanocrystal Doping by Self-Driving Fluidic Micro-Processors. Advanced Intelligent Systems *4*, 2200017. 10.1002/aisy.202200017.

34. Epps, R.W., Felton, K.C., Coley, C.W., and Abolhasani, M. (2017). Automated microfluidic platform for systematic studies of colloidal perovskite nanocrystals: Towards continuous nano-manufacturing. Lab on a Chip *17*, 4040–4047. 10.1039/C7LC00884H.

35. Li, Z., Najeeb, M.A., Alves, L., Sherman, A.Z., Shekar, V., Cruz Parrilla, P., Pendleton, I.M., Wang, W., Nega, P.W., Zeller, M., et al. (2020). Robot-Accelerated Perovskite Investigation and Discovery. Chemistry of Materials *32*, 5650–5663. 10.1021/acs.chemmater.0c01153.

36. Moradi, S., Kundu, S., Rezazadeh, M., Yeddu, V., Voznyy, O., and Saidaminov, M.I. (2022). High-throughput exploration of halide perovskite compositionally-graded films and degradation mechanisms. Communications Materials *3*, 13. 10.1038/s43246-022-00235-5.

37. Zhao, Y., Zhang, J., Xu, Z., Sun, S., Langner, S., Hartono, N.T.P., Heumueller, T., Hou, Y., Elia, J., Li, N., et al. (2021). Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning. Nature Communications *12*, 2191. 10.1038/s41467-021-22472-x.

38. Adamatzky, A. (2010). Physarum Machines: Computers from Slime Mould (World Scientific).

39. Stepney, S. (2012). Programming Unconventional Computers: Dynamics, Development, Self-Reference. Entropy [electronic only] *10*. 10.3390/e14101939.

40. Miller, J., Harding, S., and Tufte, G. (2014). Evolution-in-materio: Evolving computation in materials. Evolutionary Intelligence *7*. 10.1007/s12065-014-0106-6.

41. Schranghamer, T.F., Oberoi, A., and Das, S. (2020). Graphene memristive synapses for high precision neuromorphic computing. Nature Communications *11*, 5474. 10.1038/s41467-020-19203-z.

42. Broersma, H.J., Gomez, F., Miller, J., Petty, M., and Tufte, G. (2012). Nascence project: Nanoscale engineering for novel computation using evolution. International journal of unconventional computing *8*, 313–317.

43. Stepney, S., and Hoverd, T. (2011). Reflecting on open-ended evolution. In ECAL.

44. Ruiz-Euler, H.-C., Alegre-Ibarra, U., Ven, B., Broersma, H., Bobbert, P., and Wiel, W. (2021). Dopant network processing units: towards efficient neural network emulators with high-capacity nanoelectronic nodes. Neuromorphic Computing and Engineering *1*, 024002. 10.48550/arXiv.2007.12371,

45. Salley, D., Keenan, G., Grizou, J., Sharma, A., Martín, S., and Cronin, L. (2020). A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles. Nat Commun *11*, 1–7. 10.1038/s41467-020-16501-4.

List of Figures

List of Tables

Term	Description	Examples
Workflow	Entire sequence of *states* joined by a series of *tasks*	–
State	Things that can be acted upon, which we tend to keep records of or collect data from. Subclasses of *states* can be defined with increasing specificity	Precursors, reagents, data, models Liquid ⊃ solvent ⊃ reagent
Article	An instance of *state* class; an object with values assigned to all its attributes	1M Ammonium chloride in methanol, Data table of XRD measurements of titanium dioxide powder
Task	Acts that we perform on the *state*, which can be described in a tool-agnostic way (4 categories of *tasks*: materials synthesis, materials processing, characterisation, analysis)	Polymerization (materials synthesis) Thin film deposition (materials processing) Electrical measurement (characterisation) Data wrangling, Machine learning (analysis)
Technique	A subclass of *task* which is more specific in the method to be used, while still remaining tool-agnostic (We also acknowledge that some *techniques* naturally limit themselves to the use of a particular *tool*)	Spin coating, PVD, CVD (thin film deposition), FET measurement, 4-point probe measurement, Hall effect measurement (electrical measurement) Batch reactors, flow reactors (Polymerization) GPC measurements Transfer Learning, Ensemble Methods (ML Models)
Tool	An instance of *technique* subclass; An equipment that carries out a series of *actions* to complete a *task*	Laurell WS-650-23B spin coater, OT-2 Opentrons machine, Syringe pumps, Fraction collector, Mixers, Hotplate stirrer
Action	A method of the *tool* object; An actionable move that may be specified by the use of certain *tools*	Transfer liquid, Stir mixture, Spin substrate

Table 1: Summary of the terms used in our framework and their definition.