

From materials discovery to system optimization by integrating combinatorial electrochemistry and data science

Helge S. Stein^{1,2}, Alexey Sanin^{1,2}, Fuzhan Rahmanian^{1,2},
Bojing Zhang^{1,2}, Monika Vogler^{1,2}, Jackson K. Flowers^{1,2},
Leon Fischer^{1,2}, Stefan Fuchs^{1,2}, Nirmal Choudhary^{1,2} and
Lisa Schroeder^{1,2}

Abstract

Insight generation from electrochemical experiments augmented by data science requires broad, systematic, and well-defined parameter variations which build upon automation, data management, and flexible instrumentation interfaces. Combinatorial electrochemical synthesis of interfaces and interphases with liquid electrolytes by automated high-throughput robots offers the required high reproducibility. However, automation of electrochemistry is not enough as data needs to be collected in ways that make it machine readable and interpretable. Once established this integration allows scientists and algorithms to transfer knowledge and insights from interfaces and interphases to systems like batteries. Herein, we present an overview of recent innovative methods of combinatorial electrochemistry and synthesis which have been integrated into our platform for accelerated electrochemical storage research (PLACES/R), targeting the entire battery research value chain.

Addresses

¹ Institute of Physical Chemistry (IPC), Karlsruhe Institute of Technology (KIT), Fritz-Haber Weg 2, 76131 Karlsruhe, Germany

² Helmholtz Institute Ulm (HIU), Helmholtzstr. 11, 89081 Ulm, Germany

Corresponding author: Stein, Helge S. (helge.stein@kit.edu)

Keywords

Combinatorial, High-throughput, Electrochemistry, Data science, Machine learning, Integration, Batteries.

Introduction

Electrochemistry for energy storage or energy conversion often involves finding better materials or processes to optimize functional performance or, through well-defined parameter variation, gain a mechanistic understanding of the underlying physicochemical parameters [1]. Typically, these variations necessitate a series of (complex), often manually performed experiments [2]. Manual experimentation is time and cost-intensive, bears the problem of low reproducibility, and complicates digitalization [3,4]. A lack of data management also confounds integration [5] with other methods for the generation of composition–structure–property maps [6] and data science [7] methods. Electrochemistry performed at greater throughput, productivity, reproducibility, and with data management enables transfer learning [8] linking interfacial to system properties [9]. Such scale bridging would however not just require innovative high-throughput instrumentation and combinatorial synthesis, but also the integration between them. The integration of combinatorial synthesis, high-throughput electrochemistry, and data science is going to be the major driving force in establishing materials acceleration platforms [10] for rapid discovery, characterization, understanding and upscaling of catalysts, battery materials and electrochemical energy harvesting. In the following sections we will give brief state of the art examples and challenges for room-temperature liquid electrolyte electrochemistry and discuss how we envision surmounting these into a platform for accelerated electrochemical energy-storage research (PLACES/R).

Terminology of combinatorial chemistry

Combinatorial materials science (CMS) emerged in the early 2000s [2,11,12], with a broader adoption globally in the framework of the US lead materials genome initiative (MGI) [13,14], and formed the base for recent

efforts in the European Union and Canada to build internationally connected materials acceleration platforms (MAP) [1,5,10].

Despite the maturity of the field, there is a very liberal use of community jargon. We recommend using the term “combinatorial” whenever a large *count of combinations* is synthesized, mixed, or varied in a library [6], which consists of logically connected entities. Methods should be referred to as being “*high-throughput*” when they allow automatic measurements on “*libraries*.” In our opinion, referring to combinatorial workflows as *screening* should be avoided as the word carries the undertone of being superficial and of low fidelity.

Combinatorial electrochemistry

New innovative electrochemical methods developed within the framework of CMS and MGI include electrochemical imaging and measuring techniques like scanning electrochemical microscopy [15] (SECM) and scanning droplet cells (SDC) [16] [–] [21]. To probe electrolyte parameters, recent advances in high-throughput electrolyte conductivity measurements and optimization thereof are reported by Dave et al. [22] that utilize a series of valves and pumps connected to a viscometer and potentiostat to perform on-line conductivity measurements.

Building upon the idea of a translating electrochemical probe from SECM, several designs of SDCs have emerged. Originally used in electrocatalysis [21,23,24], SDCs were recently reviewed for their applicability in battery research by Daboss et al. [15]. SDCs solve two of the main challenges in electrochemistry: 1) rapid assembly of an electrochemical half-cell and 2) the efficient use of electrode material. This is of utmost importance especially in high-throughput catalysis [25] and battery research [15] due to the scarcity of material. The general concept of deploying SDCs in battery research is that the full cell is split into the respective half-cells in which the materials (electrodes, electrolytes) are studied individually. The major benefit of this setup is that it allows for a rapid sequential assessment of interfaces and the rapid synthesis of interphases as schematically shown in Figure 1. Only a small area of the working electrode is exposed to the electrolyte demanding about centimeter sized electrodes. This allows for a combinatorial variation in the electrode (e.g. composition, structure, processing, morphology), the electrolyte (e.g. its formulation or controlled addition of impurities), and the electrochemical protocol. Subsequent to the electrochemical experiment, high-throughput spectroscopic measurements can be performed to correlate electrochemical variations to structural ones as shown in Figure 1b.

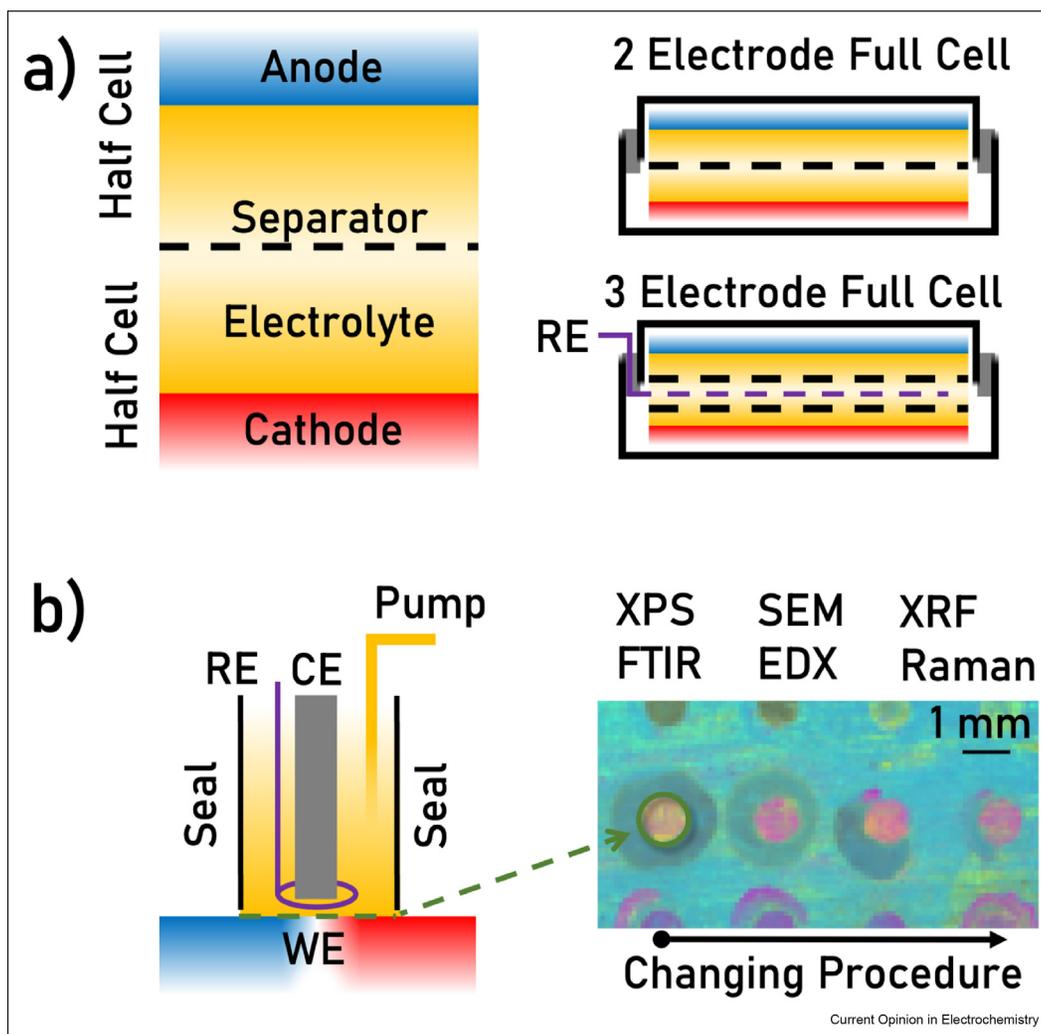
Only a few examples exist where a no flow (NF) type SDC has been used in non-aqueous battery research [26]. This, in our opinion, relates mostly to the limited choice of long-term stable and easy to fabricate non-aqueous micro-reference electrodes, and the challenge of ensuring good droplet reproducibility with low viscosity and incompatible surface tension materials like commonly used carbonate and fluorinated salt formulations in battery research. Our solution to these challenges is the simultaneous use of PTFE knife edge [15,21,23] or O-ring FFKM seals with active pressure control and on-contact cell filling with post-measurement suction.

SDC-type measurements are however designed for rapid sequential experiments leading to the fact that the overall experimental time is determined by each individual measurement’s duration. As experiments in battery research may last for up to months, other methods with parallel designs are necessary. A notable example is the use of individually addressable cells on which electrodes are sputtered as demonstrated by Fleischauer et al. [27]. This setup does however necessitate the use of a sputtering machine and relatively advanced substrate patterning. Other notable examples of combinatorial synthesis for electrochemistry include the deposition of combinatorial composition spread type libraries [28] to study not just alloying effects but also morphology [29] and processing variations [30]. Combinatorial synthesis methods also include inkjet [31,32] printing of ultra large libraries [33,34].

A need for lab-scale manufacturing

Linking materials discoveries at low technology readiness levels (TRLs) and production engineering necessitates lab-scale large batch testing of electrode, electrolyte, and protocol variations, which to date does not exist. Such workflows could help in the assessment of scalability, cell-to-cell variation, and manufacturability at an early materials discovery stage. Discoveries of new and improved materials should anyhow be considered in the light of recent publications [35] on the minimum number of required experiments to account for cell-to-cell variation. As an example, even a 10%-step variation in a ternary system (i.e. three solvents; 66 combinations) would necessitate about 594–858 cells (9–13 repeats each [35]) to exclude potentially misleading cell-to-cell variations. Comparing these numbers to the largest publicly accessible datasets on (commercial grade quality) cell testing that range from 22 to 124 (commercial) cells [35] highlights the necessity for lab scale battery production instruments, but also demands a broader adoption of FAIR (findable, accessible, interoperable, reusable) [4,36] publishing standards in electrochemistry.

Figure 1



Schematic view of cell concepts and uses of scanning droplet cells (SDC) in particular (a) full cell or secondary battery consisting of two half cells with a separator between electrodes (left side). The full-cell design is most commonly used in coin and pouch cells (right side) that are most amenable for long-term measurements. Coin cell components are expensive, not reusable, and cumbersome to disassemble for post-mortem studies. (b) SDCs are miniaturized half-cells that only expose small areas of the electrode material to the electrolyte. The SDC design is most amenable for post-mortem studies by XPS, XRF, FTIR, Raman, SEM, and EDX. The image on the right shows a false color Raman image of Cu-oxides on a Cu-foil from chronoamperometric SDC experiments. The sequential nature of SDC experiments practically limits their individual duration up to a few hours.

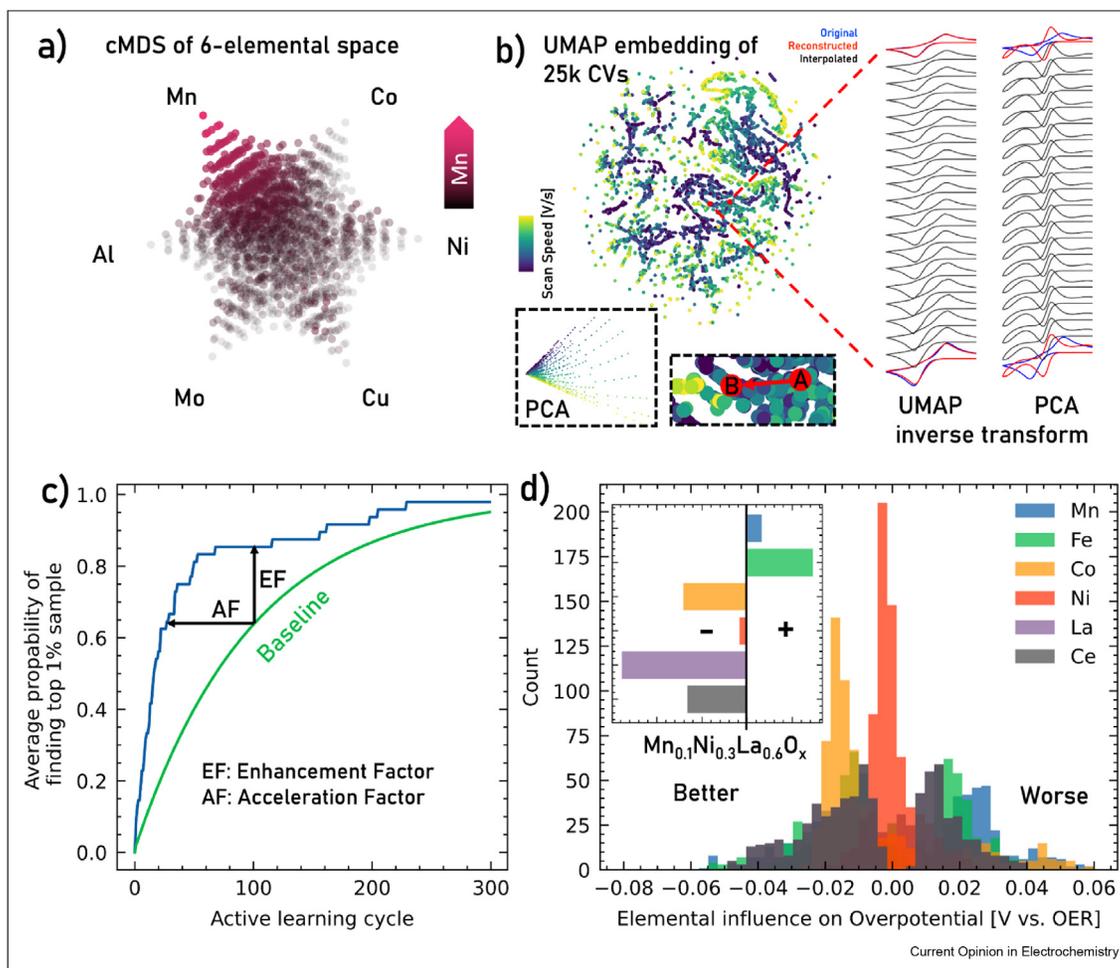
Acceleration through machine learning

Combinatorial libraries are highly amenable to data science such as machine learning (ML) once data management has been put into place. However, ML has caught on arguably late in electrochemistry, likely due to the only recent development of data lineage tracking frameworks [4,37–39]. There are a multitude of research questions that benefit from, or require the use of, ML methods. One early deployment of ML for combinatorial chemistry was the analysis of X-ray diffraction data by Takeuchi et al. [40], which sparked a great number of publications, ultimately leading to autonomous analysis agents for phase region identification [41]. We identify five emblematic tracks of applying

ML in combinatorial electrochemistry: 1) data curation, 2) data visualization, 3) data analysis, 4) active learning, and 5) automated insight generation, for which we will briefly list some examples.

Often neglected is the need for data curation, that is, background removal, denoising, and peak identification. Ament et al. [42] demonstrate the benefit of considering a full dataset to infer backgrounds, which we believe holds great potential for analyzing redox peaks in cycling voltammetry (CV) measurements [57–60] to enhance the downstream visualization by methods like manifold learning [43]. Alternate methods for noise and dimensionality reduction are autoencoder neural networks [44]

Figure 2



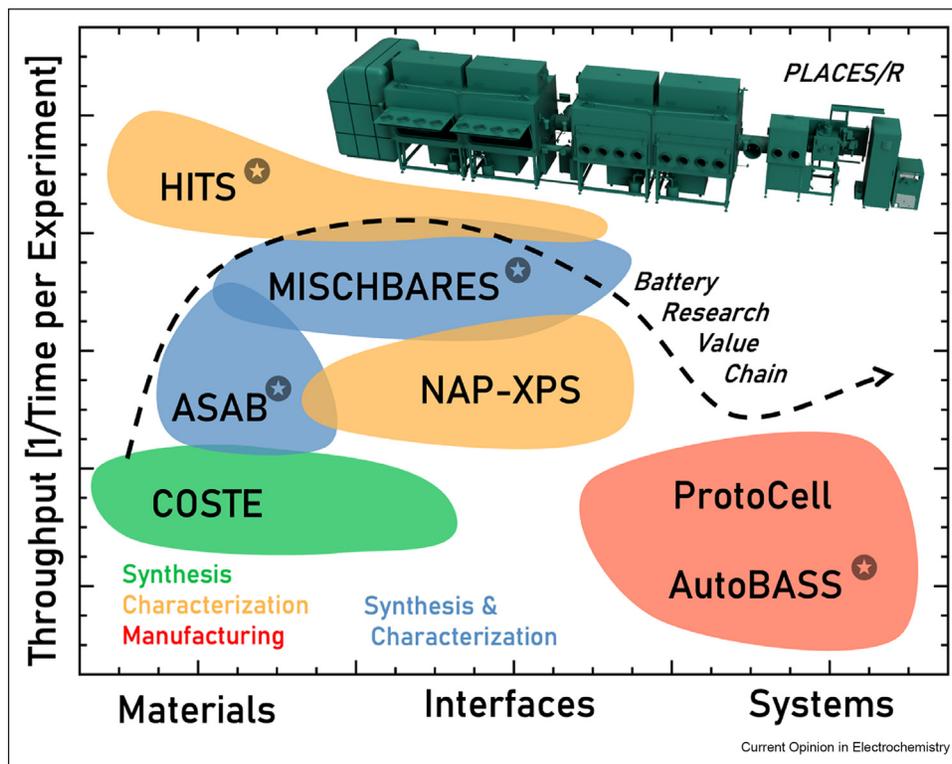
Exemplary applications of machine learning (ML) to electrochemical research questions beyond feed forward regression on OER catalysts and cyclic voltammetry: (a) compositional multidimensional scaling (cMDS) [1–3] to visualize compositional trends in a 6-component space, here highlighting the content of Mn; (b) latent space representation of 25,000 computer generated CV redox curves using different parameters comparing UMAP [48] and principal component analysis (PCA) showing the need to describe the non-linear relationships and latent space interpolation capabilities; (c) an example of active learning on OER data [25,49]; (d) an example of ML assisted generation of insights on a OER dataset using SHAP [50] coefficients allowing for the retrieval individual elemental influence factors.

or compositional multidimensional scaling (cMDS) [25]. The latter is a facile method to visualize compositional spaces of more than four components [25] as demonstrated in Figure 2a for a six elements. Other methods such as UMAP [48], shown in Figure 2b, allow not just for the reduction of data dimensionality for visualization purposes but also for interpolation in the latent space better than by traditional methods like PCA.

Most deployments of ML for data analysis fall into the category of feed forward regression or classification algorithms. Examples for early lifetime prediction were published by Attia et al. [45]. Active learning aims at accelerating and enhancing the outcomes of experimental campaigns [49], as shown in Figure 2c. There are examples of active learning for battery research, including

fast charging protocol optimization [46] to extend battery lifetime and electrolyte conductivity optimization by changing the formulation ratios [22]. Understanding of the influence of certain materials parameters on the functional properties has been long sought after and has been demonstrated by Umehara et al. [47]. In that study, a neural network predicted the photoelectrochemical power with composition and Raman spectra as input and independently rediscovered the monoclinic distortion in bismuth vanadate [24] together with previously unidentified physicochemical relationships. Emergent is the field of uncertainty quantification [49] in ML, which to date has not been applied to pure electrochemistry. In Figure 2d a simple deployment of SHAP coefficients to a dataset [49] of OER catalysts is shown, which enables a rapid identification that Co and Ni containing catalysts

Figure 3



Conceptual arrangement of the available instrumentation of the platform for accelerated electrochemical energy storage research (PLACES/R) by throughput and materials-interfaces-systems scale. Systems with a star are in-house developed. The lowest throughput is found in AutoBASS, where an assembled cell may be tested for months, and the highest throughput in HITS, where spectra can be acquired within seconds. Instruments interface with Kadi4Mat [37] and are accessible through HELAO [51]. The system allows research across the entire battery research value chain, from early materials discovery to process optimization and manufacturing.

are better than Mn and Fe containing ones. There are however some studies that aim at applying machine learning to link theory and experiment in voltametric studies [57–59], for example, to identify an electrode’s reaction mechanism [60].

From high-throughput instruments to laboratory interfaces

All the above-mentioned techniques in combinatorial electrochemistry work upon the underlying principles of research task acceleration and integration as recently reviewed by Stein & Gregoire [5]. Only when instrumentation such as SDCs or cell manufacturing robots store data in formats that go beyond FAIR, that is, including also how the data was acquired on the hardware operational level linked to ontologies, does an integration of electrochemistry with other techniques become feasible. Our approach to the process of integrating research tasks and ultimately laboratories in different locations is split in two steps: 1) the development of frameworks to automate high-throughput instruments and integrate ML methods 2) the development of hardware agnostic frameworks

to integrate multiple research facilities. For the first step, we have developed HELAO [51], which can operate the majority of the research Instruments, including the various SDCs, in our and other (e.g. Gregoire lab at Caltech) laboratories. With a single middleware for high-throughput experimentation, the overhead and time investment for integration of data analysis, data management, and ultimately ML is greatly reduced. Such an integration is necessary to deploy active learning for efficient chemical space exploration.

Going beyond high-throughput experimentation and active learning in a single lab [51–53] and towards the fulfillment of the materials acceleration platform (MAP) paradigm as envisioned by BIG-MAP [1,54,55], other modes of instrument interactions are needed. We believe the next step for the realization of MAPs is the decoupling of a measurement request from the underlying hardware, that is, only having to specify what kind of electrochemical measurement is needed without telling an instrument or collaborator how they should perform it.

Conclusion

Towards accelerated electrochemical energy storage research

The development of tools and frameworks for automation of research instruments for interface and system studies, and for the acceleration of electrochemical data curation, analysis, and understanding by means of data science methods such as active learning, shows considerable promise. With the pressing need for sustainable large-scale mobile and stationary energy storage systems in everything from smartphones, eBikes, cars, and homes, a more rapid transfer of materials from low to high TRL is necessary. In our research group, we have therefore ventured to build a platform for accelerated electrochemical energy storage research (PLACES/R) which, through automation of pertinent instrumentation and the tight integration of data science methods, seeks to cover the entire battery research value chain. As shown in Figure 3, the system consists of large glovebox filled with dry, oxygen-free, inert atmosphere that house a combinatorial reactive sputtering system for thin-film electrode synthesis (COSTE), and multiple SDCs that perform millimeter scale high-throughput battery research [15] (MISCHBARES) on thin-films and conventional electrodes. Synthesized interphases can be analyzed by a high-throughput spectroscopy (HITS) setup in which three measurement heads for Raman and FTIR are translated over a substrate. Operando interfacial studies are made possible with a near-ambient pressure XPS (NAP-XPS). Electrolytes can be studied independently of electrodes (i.e. without the influence of binders or geometries) by a setup for autonomous synthesis and analysis of battery electrolytes (ASAB) consisting of a syringe pump setup, a densimeter, a viscometer, and a broadband pulsed NMR. All these systems are employed in the early stages of materials discovery for batteries to down-select materials for upscaling in the automatic battery assembly system (AutoBASS) that can build 64-coin-cell batches and the ProtoCell robot to manufacture 120-pouch-cell batches. In addition to the hardware, all instruments can interface to the central data management platform Kadi4Mat [37] and can be operated through HELAO [51]. All mentioned instruments will be published elsewhere in due time.

Through HELAO and Kadi4Mat, we can accelerate instruments using active learning but also construct datasets more rapidly as we have complete data lineage tracking. In a roadmap towards the establishment of a truly integrated materials acceleration platform [1] we see the current integration of ML for optimizers and automatic data analysis as the first step. What needs to follow, is a hardware independent API to our laboratory enabling other human or robotic scientists to conduct research in our lab by only telling us *what* they want without telling us necessarily *how* we're supposed to acquire the information. In our

opinion this serves as the foundation to address the Nobel-Turing-Challenge [56].

We believe that the discovery, upscaling, and insight generation of sustainable and high-performance batteries can be greatly accelerated through the integration of combinatorial synthesis, high-throughput electrochemistry, automatic battery assembly, and a centralized orchestration framework with ML augmentation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Acknowledgements

This work contributes to the research performed at CELEST (Center for Electrochemical Energy Storage Ulm-Karlsruhe) and was funded by the German Research Foundation (DFG) under Project ID 390874152 (POLiS Cluster of Excellence). This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189. This project received funding from Bundesministerium für Bildung und Forschung (BMBF) in the framework of the BMBF-Kompetenzcluster InZePro for the projects DataBatt No 03XP0323D and InForm No 03XP0363A.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- * of special interest
- ** of outstanding interest

1. Amici J, Asinari P, Ayerbe E, Barboux P, Bayle-Guillemaud P, Behm RJ, Berecibar M, Berg E, Bhowmik A, Bodoardo S, Castelli IE, Cekic-Laskovic I, Christensen R, Clark S, Diehm R, Dominko R, Fichtner M, Franco AA, Grimaud A, Guillet N, Hahlin M, Hartmann S, Heiries V, Hermansson K, Heuer A, Jana S, Jabbour L, Kallo J, Latz A, Lormann H, Løvvik OM, Lyonnard S, Meeus M, Paillard E, Perraud S, Placke T, Punckt C, Raccurt O, Ruhland J, Sheridan E, Stein H, Tarascon J, Trapp V, Vegge T, Weil M, Wenzel W, Winter M, Wolf A, Edström K: **A roadmap for transforming research to invent the batteries of the future designed within the European large scale research initiative BATTERY 2030+.** *Adv Energy Mater* 2022: 2102785, <https://doi.org/10.1002/aenm.202102785>.
The European Battery research community presents the current state of the art and a way forward in how they envision to establish a collaborative battery research ecosystem under the umbrella of Batter 2030+. This is of utmost importance for the current and future strategy of electrochemical energy storage.
2. Potyrailo R, Rajan K, Stoewe K, Takeuchi I, Chisholm B, Lam H: **Combinatorial and high-throughput screening of materials libraries: review of state of the art.** *ACS Comb Sci* 2011, **13**: 579–633, <https://doi.org/10.1021/co200007w>.
3. Banko L, Ludwig A: **Fast-track to research data management in experimental material science—setting the ground for research group level materials digitalization.** *ACS Comb Sci* 2020, **22**:401–409, <https://doi.org/10.1021/acscombsci.0c00057>.
Without data management there cannot be data science and herein a methodology is presented in which document management systems are used to track samples, analyses and insights to remove bottlenecks in the overall research process. This is important as electrochemistry is never a means in itself and relies on metadata of what is being measured. This manuscript present a way to accelerate that process.
4. Castelli IE, Arismendi-Arrieta DJ, Bhowmik A, Cekic-Laskovic I, Clark S, Dominko R, Flores E, Flowers J, Frederiksen KU, Friis J, Grimaud A, Hansen KV, Hardwick LJ, Hermansson K, Königler L, Lauritzen H, Cras FL, Li H, Lyonnard S, Lormann H, Marzari N,

- Niedzicki L, Pizzi G, Rahmanian F, Stein H, Uhrin M, Wenzel W, Winter M, Wölke C, Vegge T: *Data management plans: the importance of data management in the BIG-MAP project*. 2021. ArXiv210601616 Cond-Mat.
- Stein HS, Gregoire JM: **Progress and prospects for accelerating materials science with automated and autonomous workflows**. *Chem Sci* 2019, **10**:9640–9649, <https://doi.org/10.1039/C9SC03766G>.
 - Ludwig A: **Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods**. *NPJ Comput Mater* 2019, **5**, <https://doi.org/10.1038/s41524-019-0205-0>.
 - Agrawal A, Choudhary A: *Perspective: materials informatics and big data: realization of the “fourth paradigm” of science in materials science*, vol. 11; 2016.
 - Kong, S.; Guevarra, D.; Gomes, C. P.; Gregoire, J. M. Materials representation and transfer learning for multi-property prediction. vol. 14.
 - Fichtner M: **Recent research and progress in batteries for electric vehicles**. *Batter Supercaps* 2022, **5**, <https://doi.org/10.1002/batt.202100224>.
A review on the drastic price reduction of 98% and capacity increase of 3x in the past three decades and attributes these findings mostly to engineering and less to materials improvements. This perspective on materials selection in actually utilized batteries opens up new avenues for materials selection and design.
 - Materials acceleration platform—accelerating advanced energy materials discovery by integrating high-throughput methods with artificial intelligence. vol. 109.
 - Amis EJ, Xiang XD, Zhao JC: **Combinatorial materials science: what’s new since edison?** *MRS Bull* 2002, **27**:295–300.
 - Maier WF: **Early years of high-throughput experimentation and combinatorial approaches in catalysis and materials science**. *ACS Comb Sci* 2019, **21**:437–444, <https://doi.org/10.1021/acscmbosci.8b00189>.
 - Green ML, Choi CL, Hatrick-Simpers JR, Joshi AM, Takeuchi I, Barron SC, Campo E, Chiang T, Empedocles S, Gregoire JM, Kusne AG, Martin J, Mehta A, Persson K, Trautt Z, Van Duren J, Zakutayev A: **Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies**. *Appl Phys Rev* 2017, **4**:11105, <https://doi.org/10.1063/1.4977487>.
 - Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson KA: **Commentary: the materials project: a materials genome approach to accelerating materials innovation**. *Apl Mater* 2013, **1**, 011002.
 - Daboss, S.; Rahmanian, F.; Stein, H. S.; Kranz, C. The Potential of Scanning Electrochemical Probe Microscopy and Scanning Droplet Cells in Battery Research. *Electrochem Sci Adv n/a (n/a)*, e2100122. <https://doi.org/10.1002/elsa.202100122>.
 - Gregoire JM, Xiang C, Liu X, Marcin M, Jin J: **Scanning droplet cell for high throughput electrochemical and photo-electrochemical measurements**. *Rev Sci Instrum* 2013, **84**, 024102.
 - Kollender JP, Mardare AI, Hassel AW: **Multi-scanning droplet cell microscopy (Multi-SDCM) for truly parallel high throughput electrochemical experimentation**. *Electrochim Acta* 2015, **179**:32–37, <https://doi.org/10.1016/j.electacta.2015.04.103>.
 - Grote JP, Zeradjanin AR, Cherevko S, Mayrhofer KJJ: **Coupling of a scanning flow cell with online electrochemical mass spectrometry for screening of reaction selectivity**. *Rev Sci Instrum* 2014, **85**:104101, <https://doi.org/10.1063/1.4896755>.
 - Jin J, Gregoire JM, Xiang C: *Scanning drop sensor*. December 2013.
 - Mardare AI, Ludwig A, Savan A, Hassel AW: **Scanning droplet cell microscopy on a wide range hafnium–niobium thin film combinatorial library**. *Electrochim Acta* 2013, **110**:539–549, <https://doi.org/10.1016/j.electacta.2013.03.065>.
 - Sliozberg K, Schäfer D, Erichsen T, Meyer R, Khare C, Ludwig A, Schuhmann W: **High-throughput screening of thin-film semiconductor material libraries I: system development and case study for Ti-W-O**. *ChemSusChem* 2015, **8**:1270–1278, <https://doi.org/10.1002/cssc.201402917>.
 - Dave A, Mitchell J, Burke S, Lin H, Whitacre J, Viswanathan V: *Autonomous optimization of nonaqueous battery electrolytes via robotic experimentation and machine learning*. 2021. ArXiv211114786 Cs.
 - Meyer R, Sliozberg K, Khare C, Schuhmann W, Ludwig A: **High-throughput screening of thin-film semiconductor material libraries II: characterization of Fe-W-O libraries**. *ChemSusChem* 2015, **8**:1279–1285, <https://doi.org/10.1002/cssc.201402918>.
 - Newhouse PF, Guevarra D, Umehara M, Reyes-Lillo SE, Zhou L, Boyd DA, Suram SK, Cooper JK, Haber JA, Neaton JB, Gregoire JM: **Combinatorial alloying improves bismuth vanadate photoanodes via reduced monoclinic distortion**. *Energy Environ Sci* 2018, **11**:2444–2457, <https://doi.org/10.1039/C8EE00179K>.
 - Stein H, Guevarra D, Shinde A, Jones RJR, Gregoire J, Haber JA: **Functional mapping reveals mechanistic clusters for OER catalysis across (Cu-Mn-Ta-Co-Sn-Fe)ox composition and PH space**. *Mater Horiz* 2019, **6**:1251, <https://doi.org/10.1039/C8MH01641K>.
 - Garcia G, Ventosa E, Schuhmann W: **Complete prevention of dendrite formation in Zn metal anodes by means of pulsed charging protocols**. *ACS Appl Mater Interfaces* 2017, **9**:18691–18698, <https://doi.org/10.1021/acsami.7b01705>.
 - Fleischauer MD, Hatchard TD, Bonakdarpour A, Dahn JR: **Combinatorial investigations of advanced Li-ion rechargeable battery electrode materials**. *Meas Sci Technol* 2005, **16**:212–220, <https://doi.org/10.1088/0957-0233/16/1/028>.
 - Dahn JR, Trussler S, Hatchard TD, Bonakdarpour A, Mueller-Neuhaus JR, Hewitt KC, Fleischauer M: **Economical sputtering system to produce large-size composition-spread libraries having linear and orthogonal stoichiometry variations**. *Chem Mater* 2002, **14**:3519–3523, <https://doi.org/10.1021/cm020236x>.
 - Borhani-Haghighi S, Khare C, Trócoli R, Dushina A, Kieschnick Li, LaMantia F, Ludwig A: **Synthesis of nano-structured LiMn 2 O 4 thin films by glancing angle deposition for Li-ion battery applications**. *Nanotechnology* 2016, **27**:455402, <https://doi.org/10.1088/0957-4484/27/45/455402>.
 - Stein H, Naujoks D, Grochla D, Khare C, Gutkowski R, Grütze S, Schuhmann W, Ludwig A: **A structure zone diagram obtained by simultaneous deposition on a novel step heater: a case study for Cu2O thin films**. *Phys Status Solidi Appl Mater Sci* 2015, **212**, <https://doi.org/10.1002/pssa.201532384>.
 - Liu X, Shen Y, Yang R, Zou S, Ji X, Shi L, Zhang Y, Liu D, Xiao L, Zheng X, Li S, Fan J, Stucky GD: **Inkjet printing assisted synthesis of multicomponent mesoporous metal oxides for ultrafast catalyst exploration**. *Nano Lett* 2012, **12**:5733–5739, <https://doi.org/10.1021/nl302992q>.
 - Newhouse PF, Parkinson BA: **Combinatorial optimization of spinel Co 3\textminusxM XO 4M = (Al, Ga, in) alloyed thin films prepared by ink jet printing: photoelectrochemical, optical, and structural properties**. *J Mater Chem* 2015, **3**:5901–5907, <https://doi.org/10.1039/C4TA05671J>.
 - Stein HS, Soedarmadji E, Newhouse PF, Guevarra Dan, Gregoire JM: **Synthesis, optical imaging, and absorption spectroscopy data for 179072 metal oxides**. *Sci Data* 2019, **6**:9, <https://doi.org/10.1038/s41597-019-0019-4>.
 - Yang L, Haber JA, Armstrong Z, Yang SJ, Kan K, Zhou L, Richter MH, Roat C, Wagner N, Coram M, Berndt M, Riley P, Gregoire JM: **Discovery of complex oxides via automated experiments and data science**. *Proc Natl Acad Sci Unit States Am* 2021, **118**, e2106042118, <https://doi.org/10.1073/pnas.2106042118>.
 - Dechent P, Greenbank S, Hildenbrand F, Jbabdi S, Sauer DU, Howey PA: **Estimation of Li-ion degradation test sample sizes required to understand cell-to-cell variability**. *Batter Supercaps* 2021, **4**:1821–1829, <https://doi.org/10.1002/batt.202100148>.

36. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B: **The FAIR guiding principles for scientific data management and stewardship**. *Sci Data* 2016, **3**: 160018–160019, <https://doi.org/10.1038/sdata.2016.18>.
37. Brandt N, Griem L, Herrmann C, Schoof E, Tosato G, Zhao Y, Zschumme P, Selzer M: **Kadi4Mat: a research data infrastructure for materials science**. *Data Sci J* 2021, **20**:8, <https://doi.org/10.5334/dsj-2021-008>.
- How can research data management be deployed to an entire university and beyond; this manuscript reports the Karlsruhe Data Infrastructure for materials that is the foundation to many large scale research projects. This deployment of research data management infrastructures on an institutional level is important as it opens up new research directions and collaborations.
38. Soedarmadji E, Stein HS, Suram SK, Guevarra D, Gregoire JM: **Tracking materials science data lineage to manage millions of materials experiments and analyses**. *NPJ Comput Mater* 2019, **5**:79, <https://doi.org/10.1038/s41524-019-0216-x>.
39. Zakutayev A, Perkins J, Schwarting M, White R, Munch K, Tumas W, Wunder N, Phillips CO: **High throughput experimental materials database**. 2017, <https://doi.org/10.7799/1407128>.
40. Takeuchi I, Long CJ, Famodu OO: **Data management and visualization of X-ray diffraction spectra from thin film ternary composition spreads**. *Rev Sci Ldots* 2005, <https://doi.org/10.1063/1.1927079%20>.
41. Gomes, C. P.; Bai, J.; Xue, Y.; Björck, J.; Rappazzo, B.; Ament, S.; Bernstein, R.; Kong, S.; Suram, S. K.; Dover, R. B. van; Gregoire, J. M. CRYSTAL: A Multi-Agent AI System for Automated Mapping of Materials' Crystal Structures. *MRS Commun.* undefined/ed, 1–9. <https://doi.org/10.1557/mrc.2019.50>.
42. Ament SE, Stein HS, Guevarra D, Zhou L, Haber JA, Boyd DA, Umehara M, Gregoire JM, Gomes CP: **Multi-component background learning automates signal detection for spectroscopic data**. *NPJ Comput Mater* 2019, **5**:1–7, <https://doi.org/10.1038/s41524-019-0213-0>.
43. Vaddi K, Wodo O: **Metric learning for high-throughput combinatorial data sets**. *ACS Comb Sci* 2019, <https://doi.org/10.1021/acscombsci.9b00086>.
44. Stein HS, Guevarra D, Newhouse PF, Soedarmadji E, Gregoire JM: **Machine learning of optical properties of materials – predicting spectra from images and images from spectra**. *Chem Sci* 2019, **10**:47–55, <https://doi.org/10.1039/C8SC03077D>.
45. Herring P, Balaji Gopal C, Aykol M, Montoya JH, Anapolsky A, Attia PM, Gent W, Hummelshøj JS, Hung L, Kwon H-K, Moore P, Schweigert D, Severson KA, Suram S, Yang Z, Braatz RD, Storey BD: **BEEP: a Python library for battery evaluation and early prediction**. *Software* 2020, **11**:100506, <https://doi.org/10.1016/j.softx.2020.100506>.
- Data analysis in battery research and early lifetime prediction are streamlined in a easy to use software package that will allow researchers to fast track their research without having to reinvent the wheel.
46. Attia PM, Grover A, Jin N, Severson KA, Markov TM, Liao Y-H, Chen MH, Cheong B, Perkins N, Yang Z, Herring PK, Aykol M, Harris SJ, Braatz RD, Ermon S, Chueh WC: **Closed-loop optimization of fast-charging protocols for batteries with machine learning**. *Nature* 2020, **578**:397–402, <https://doi.org/10.1038/s41586-020-1994-5>.
47. Umehara M, Stein HS, Guevarra D, Newhouse PF, Boyd DA, Gregoire JM: **Analyzing machine learning models to accelerate generation of fundamental materials insights**. *NPJ Comput Mater* 2019, **5**, <https://doi.org/10.1038/s41524-019-0172-5>.
48. Sainburg T, McInnes L, Gentner TQ: **Parametric UMAP embeddings for representation and semi-supervised learning**. 2021. [ArXiv200912981](https://arxiv.org/abs/2009.12981) Cs Q-Bio Stat.
49. Rohr B, Stein HS, Guevarra D, Wang Y, Haber JA, Aykol M, Suram SK, Gregoire JM: **Benchmarking the acceleration of materials discovery by sequential learning**. *Chem Sci* 2020, **11**:2696–2706, <https://doi.org/10.1039/C9SC05999G>.
- Active learning is a process in which acquired data is utilized to suggest the optimal next experiment in an autonomous fashion to improve a model and the outcome. This study presents benchmarks and a demonstration on oxygen evolution reaction catalysts and how research can be accelerated by up to 10x
50. Lundberg SM, Lee S-I: **A unified approach to interpreting model predictions**. In *Advances in Neural Information Processing Systems 30*. Edited by Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, Curran Associates, Inc.; 2017:4765–4774.
51. Rahmanian F, Flowers J, Guevarra D, Richter M, Fichtner M, Donnelly P, Gregoire JM, Stein HS: **Enabling modular autonomous feedback-loops in materials science through hierarchical experimental laboratory automation and orchestration**. *Adv Mater Interfac* 2022:2101987, <https://doi.org/10.1002/admi.202101987>.
52. Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LPE, Hein JE, Aspuru-Guzik A: **ChemOS: orchestrating autonomous experimentation**. *Sci Robot* 2018, **3**, <https://doi.org/10.1126/scirobotics.aat5559>.
53. Allan D, Caswell T, Campbell S, Rakitin M: **Bluesky's ahead: a multi-facility collaboration for an a La carte software project for data acquisition and management**. *Synchrotron Radiat News* 2019, **32**:19–22, <https://doi.org/10.1080/08940886.2019.1608121>.
54. Bhowmik A, Berecibar M, Casas-Cabanas M, Csanyi G, Dominko R, Hermansson K, Palacin MR, Stein HS, Vegge T: **Implications of the BATTERY 2030+ AI-assisted toolkit on future low-TRL battery discoveries and chemistries**. *Adv Energy Mater* 2021:2102698, <https://doi.org/10.1002/aenm.202102698>.
- Building upon the data and knowledge on existing established chemistries the Battery2030+ project develops data science tools to accelerate research and herein presents a perspective on how these machine learning tools may accelerate future battery chemistries
55. Bhowmik A, Castelli IE, Garcia-Lastra JM, Jørgensen PB, Winther O, Vegge T: **A perspective on inverse design of battery interphases using multi-scale modelling, experiments and generative deep learning**. *Energy Storage Mater* 2019, **21**: 446–456, <https://doi.org/10.1016/j.ensm.2019.06.011>.
56. Kitano H: **Nobel Turing Challenge: creating the engine for scientific discovery**. *NPJ Syst Biol Appl* 2021, **7**:29, <https://doi.org/10.1038/s41540-021-00189-3>.
57. Gundry L, Kennedy G, Bond AM, Zhang J: **Inclusion of multiple cycling of potential in the deep neural network classification of voltammetric reaction mechanisms**. *Faraday Discuss* 2022, **233**:44–57, <https://doi.org/10.1039/D1FD00050K>.
58. Gundry L, Guo S-X, Kennedy G, Keith J, Robinson M, Gavaghan D, Bond AM, Zhang J: **Recent advances and future perspectives for automated parameterisation, bayesian inference and machine learning in voltammetry**. *Chem Commun* 2021, **57**:1855–1870, <https://doi.org/10.1039/D0CC07549C>.
59. Bond AM, Zhang J, Gundry L, Kennedy GF: **Opportunities and challenges in applying machine learning to voltammetric mechanistic studies**. *Curr Opin Electrochem* 2022:101009, <https://doi.org/10.1016/j.coelec.2022.101009>.
60. Kennedy GF, Zhang J, Bond AM: **Automatically identifying electrode reaction mechanisms using deep neural networks**. *Anal Chem* 2019, **91**:12220–12227, <https://doi.org/10.1021/acs.analchem.9b01891>.