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

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

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 \cite{1}. Typically, these variations necessitate a series of (complex), often manually performed experiments \cite{2}. Manual experimentation is time and cost-intensive, bears the problem of low reproducibility, and complicates digitalization \cite{3,4}. A lack of data management also confounds integration \cite{5} with other methods for the generation of composition–structure–property maps \cite{6} and data science \cite{7} methods. Electrochemistry performed at greater throughput, productivity, reproducibility, and with data management enables transfer learning \cite{8} linking interfacial to system properties \cite{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 \cite{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 \cite{2,11,12}, with a broader adoption globally in the framework of the US lead materials genome initiative (MGI) \cite{13,14}, and formed the base for recent

Addresses
\textsuperscript{1} Institute of Physical Chemistry (IPC), Karlsruhe Institute of Technology (KIT), Fritz-Haber Weg 2, 76131 Karlsruhe, Germany
\textsuperscript{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.
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.
**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 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.
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...
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 (HTS) 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 (POLAS Cluster of Excellence). This project received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957199. This project received funding from Bundesministerium für Bildung und Forschung (BMBF) in the framework of the BMBF-Kompetenzzentrum 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


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.


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/acscombscil.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.


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.


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.


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.


