An Integrated Platform for In Situ Electroanalytical–Driven Reaction Optimization

Timothy J. McClure [*1] Elvis C. McFee [*1] Shi Xuan Leong [*2,4] Seren G. Parikh [1] Luna D. Paffrath [1] Sergio Pablo-García Carrillo [3] Matthew E. Reish [1] Brandon Wong [5] Stephen Maldonado [6] Alán Aspuru-Guzik[3,4,5] Corinna S. Schindler[1]

[*]Equal contribution [1]Department of Chemistry, University of British Columbia, Vancouver V6T 1Z1 BC, Canada [2] Department of Chemistry, School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, 21 Nanyang Link, Singapore 6373713 [3]Acceleration Consortium, 700 University Ave., M7A 2S4, Toronto, Canada, [4]Department of Chemistry, University of Toronto, Lash Miller Chemical Laboratories, 80 St. George Street, ON M5S 3H6, Toronto, Canada, [5]Department of Computer Science, University of Toronto, Sandford Fleming Building, 10 King's College Road, ON M5S 3G4, Toronto, Canada, [6]Department of Chemistry and Program in Applied Physics, University of Michigan, Ann Arbor, Michigan 48109-1055, United States

Correspondence to: Corinna S. Schindler schindler@chem.ubc.ca

1. Introduction

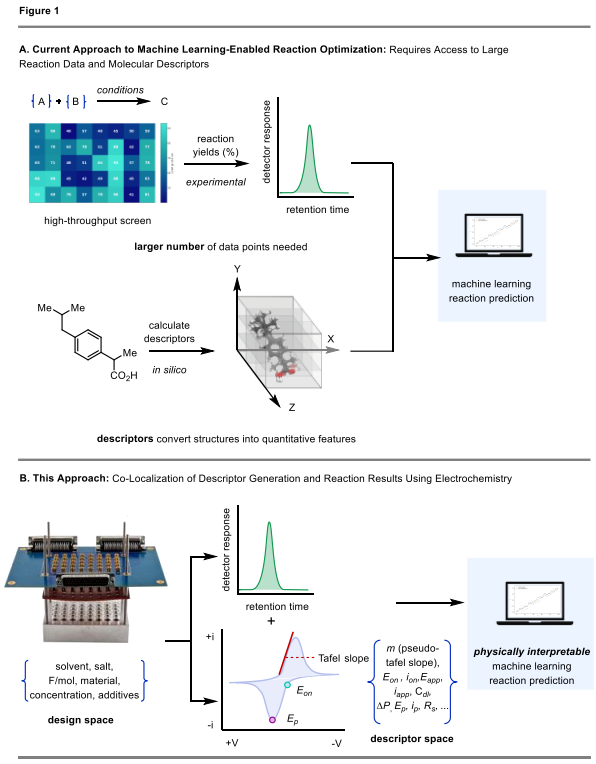

Figure 1A. Classical approach to HTE-ML, pairing DFT and exhaustive HTE. B. Co-localization of physical descriptor collection and reactions of inter est.

Machine-learning–guided modeling has substantially accelerated discovery across the sciences by enabling efficient exploration of high-dimensional experimental design spaces.[1] [2] [3] [4] [5] [6] In chemical synthesis and related domains, these advances have been most successful when models are trained on carefully curated datasets encoded with physically meaningful descriptors. However, the scalability of such approaches is increasingly constrained by a fundamental trade-off between experimental and computational cost that arises from descriptor choice. [7] [8] [9] [10] Interpretable, physics-informed descriptors often require expensive computation or extensive experimental characterization, while lightweight encodings scale efficiently but provide only indirect representations of the operative system.[11] [12] As a result, descriptor generation is commonly decoupled from the experiment itself, relying on abstractions (such as those from Density Functional Theory, DFT) rather than measurements acquired under operating conditions, which increases data requirements while decreasing closed-loop efficiency (Figure 1A). [13] Here, we introduce a measurement-native strategy in which in situ electroanalytical data is directly transformed into machine learning-ready descriptors within the experimental setup (Figure 1B). By embedding descriptor generation into experiment execution, this framework unifies measurement, modeling, and optimization in a single *loop*, enabling data-efficient acceleration without sacrificing physical grounding. While demonstrated using electrochemical synthesis, we envision the generalization of this approach to any physical system where informative measurements can be acquired in situ.

2. Main

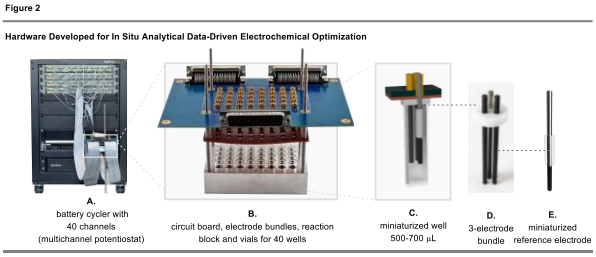

Figure 2. Hardware stack for in situ electrodescriptor collection and electrolysis.

To address limitations of existing electrochemical platforms [14], we developed and validated an integrated hardware stack for scalable, reference-standardized acquisition of in situ electroanalytical data at high throughput (Figure

2). Central to this system is a custom 40-cell, miniaturized three-electrode architecture that provides fully independent electrochemical control while co-localizing electroanalytical characterization and electrosynthesis within a shared physical environment (Figure 2B). This design preserves a stable potential reference across diverse reaction conditions, enabling reproducible descriptor generation compatible with machine-learning workflows.

A key advance is the development of a miniaturized solid-state quasi-reference electrode optimized for parallel operation (Figure 2E). Conventional junction-based references are incompatible with microscale formats, while wire pseudo-references exhibit substantial drift. We implemented a polymeric quasi-reference electrode based on electropolymerized polypyrrole on a metallic core, which displays markedly reduced potential drift under synthetically relevant conditions. [15] This enables stable, window-consistent electroanalytical measurements across extended acquisition sequences and across all 40 cells simultaneously, establishing a state-of-the-art hardware substrate for data-rich electrochemical experimentation.

The hardware platform is coupled to a custom software stack enabling automated potentiostat control, data acquisition, and analysis. A bespoke backend provides programmatic access to each channel, supports scripted electroanalytical protocols, and enables working–counter electrode inversion via a custom relay board. Electroanalytical measurements, including cyclic voltammetry and impedance spectroscopy, are thus acquired reproducibly and without manual intervention immediately prior to electrolysis for every reaction condition.

To overcome bottlenecks in electroanalytical data processing, we developed PeakYourPoint, an automated feature-extraction framework leveraging a fine-tuned vision–language model (VLM). Experimental voltammograms are rendered as standardized images with known axis mappings, allowing the VLM—trained on over 20,000 literature, experimental, and synthetic CV images—to localize redox onsets and peak features directly in image space. These features are deterministically converted into precise potential and current values, enabling fully automated extraction of physically grounded descriptors. PeakYourPoint achieves state-of-the-art performance in this context, outperforming classical signal-processing approaches, raw-signal neural networks, and general-purpose VLMs.

The integrated hardware–software stack enables closed-loop optimization using Bayesian optimization to efficiently navigate experimentally expensive, high-dimensional design spaces. We applied this framework to optimization and subsequent yield modeling of the aza-Shono reaction across a design space exceeding 100,000 parameter combinations. In situ electroanalytical descriptors acquired prior to electrolysis were used both to guide adaptive experiment selection and to train predictive yield models, demonstrating efficient optimization within experimentally tractable campaign sizes.

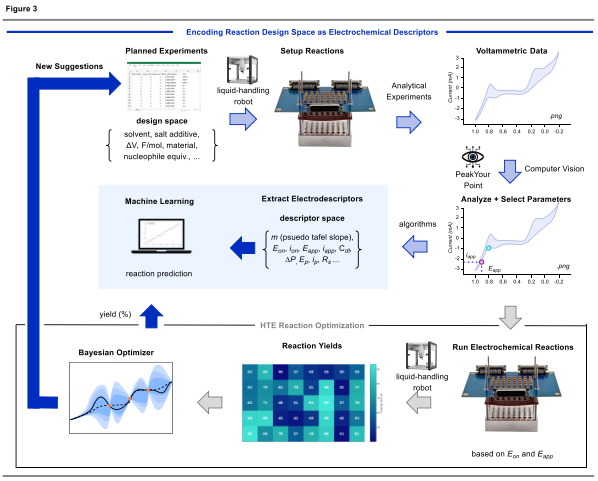

Figure 3. Experimetnal workflow for reaction optimization and reaction/ descriptor space exploration

Collectively, this work establishes a state-of-the-art experimental and computational platform in which reference-standardized hardware, automated measurement-to-descriptor pipelines, and machine-learning–guided optimization are unified within a single closed loop (Figure 3). While demonstrated for electrochemical synthesis, the underlying principles are broadly applicable to accelerated discovery across the physical sciences.

Acknowledgments

We thank the Acceleration Consortium (AC), the Canadian Foundation for Innovation (CFI), the Canadian Institutes of Health Research (CIHR), the Canada Excellent Research Chairs (CERC), and the Natural Sciences and Engineering Research Council (NSERC) for financial support. A.A.G. thanks Anders G. Frøseth for his generous support. A.A.G. also acknowledges the generous support of Natural Resources Canada and

the Canada 150 Research Chairs program. This research is part of the University of Toronto's Acceleration Consortium, which receives funding from the CFREF-2022-00042 Canada First Research Excellence Fund. S.X.L acknowledges fellowship support from Nanyang Technological University, Singapore and Ministry of Education, Singapore. In addition, the authors would like to thank Douglas Dickey (UM Electronics) for his help in designing the circuit board used herein, Vijay Parikh for discussions related to the development of our custom software infrastructure, Dr. Roman Krems (UBC) for discussions related to machine learning, and Lauren Rich (Maldonado Lab, UM) for help with laser cutting materials.

References

(1) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(2) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620* (7976), 1089–1100. https://doi.org/10.1038/s41586-023-06415-8.

(3) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling Deep Learning for Materials Discovery. *Nature* **2023**, *624* (7990), 80–85. https://doi.org/10.1038/s41586-023-06735-9.

(4) Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Milsted, D.; McDermott, M. J.; Gallant, M.; Cubuk, E. D.; Merchant, A.; Kim, H.; Jain, A.; Bartel, C. J.; Persson, K.; Zeng, Y.; Ceder, G. An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials. *Nature* **2023**, *624* (7990), 86–91. https://doi.org/10.1038/s41586-023-06734-w.

(5) Strieth-Kalthoff, F.; Hao, H.; Rathore, V.; Derasp, J.; Gaudin, T.; Angello, N. H.; Seifrid, M.; Trushina, E.; Guy, M.; Liu, J.; Tang, X.; Mamada, M.; Wang, W.; Tsagaantsooj, T.; Lavigne, C.; Pollice, R.; Wu, T. C.; Hotta, K.; Bodo, L.; Li, S.; Haddadnia, M.; Wołos, A.; Roszak, R.; Ser, C. T.; Bozal-Ginesta, C.; Hickman, R. J.; Vestfrid, J.; Aguilar-Granda, A.; Klimareva, E. L.; Sigerson, R. C.; Hou, W.; Gahler, D.; Lach, S.; Warzybok, A.; Borodin, O.; Rohrbach, S.; Sanchez-Lengeling, B.; Adachi, C.; Grzybowski, B. A.; Cronin, L.; Hein, J. E.; Burke, M. D.; Aspuru-Guzik, A. Delocalized, Asynchronous, Closed-Loop Discovery of Organic Laser Emitters. *Science* **2024**, *384* (6697), eadk9227. https://doi.org/10.1126/science.adk9227.

(6) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discov.* **2019**, *18* (6), 463–477. https://doi.org/10.1038/s41573-019-0024-5.

(7) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363* (6424), eaau5631. https://doi.org/10.1126/science.aau5631.

(8) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360* (6385), 186–190. https://doi.org/10.1126/science.aar5169.

(9) Reid, J. P.; Sigman, M. S. Holistic Prediction of Enantioselectivity in Asymmetric Catalysis. *Nature* **2019**, *571* (7765), 343–348. https://doi.org/10.1038/s41586-019-1384-z.

(10) Voinarovska, V.; Kabeshov, M.; Dudenko, D.; Genheden, S.; Tetko, I. V. When Yield Prediction Does Not Yield Prediction: An Overview of the Current Challenges. *J. Chem. Inf. Model.* **2024**, *64* (1), 42–56. https://doi.org/10.1021/acs.jcim.3c01524.

(11) Probst, D.; Schwaller, P.; Reymond, J.-L. Reaction Classification and Yield Prediction Using the Differential Reaction Fingerprint DRFP. *Digit. Discov.* **2022**, *1* (2), 91–97. https://doi.org/10.1039/D1DD00006C.

(12) Janssen, K.; Proppe, J. Predicting and Explaining Yields with Machine Learning for Carboxylated Azoles and Beyond. *J. Chem. Inf. Model.* **2025**, *65* (4), 1862–1872. https://doi.org/10.1021/acs.jcim.4c02336.

(13) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7* (10), 1622–1637. https://doi.org/10.1021/acscentsci.1c00535.

(14) Wills, A. G.; Charvet, S.; Battilocchio, C.; Scarborough, C. C.; Wheelhouse, K. M. P.; Poole, D. L.; Carson, N.; Vantourout, J. C. High-Throughput Electrochemistry: State of the Art, Challenges, and Perspective. *Org. Process Res. Dev.* **2021**, *25* (12), 2587–2600. https://doi.org/10.1021/acs.oprd.1c00167.

(15) Ghilane, J.; Hapiot, P.; Bard, A. J. Metal/Polypyrrole Quasi-Reference Electrode for Voltammetry in Nonaqueous and Aqueous Solutions. *Anal. Chem.* **2006**, *78* (19), 6868–6872. https://doi.org/10.1021/ac060818o.