Autonomous Experimentation as a Paradigm for Materials Discovery

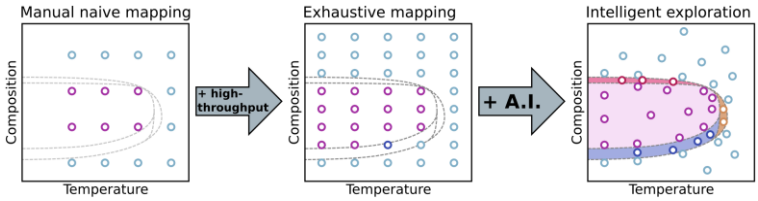
Kevin G. Yager, *Center for Functional Nanomaterials, Brookhaven National Laboratory*

Concept

One of the grand challenges of twenty-first-century materials science is the rational design of new materials, where given a desired material functionality, the material structure is predicted; and for that particular structure, we can design appropriate constituents and assembly processes. To address this challenge, it is critical to understand the relationships between constituents, processing, and resultant materials structure and function. With the needs for material functionality becoming more diverse, stringent and sophisticated, the complexity of materials continues to increase. The relevant parameter space expands correspondingly, arising from both the multi-com ponent nature of functional materials and a multitude of processing conditions. All this implies that optimizing functionality requires strategic exploration of the vast parameter space that is associated with complex materials. To meet this challenge, the way we investigate materials needs to evolve, to become more efficient and intelligent.

An emerging paradigm to address this complexity is *autonomous experimentation*, wherein experimental synthesis and data collection are automated, and machine-learning algorithms are used to select experiments to conduct based on the evolving dataset. Implemented properly, these methods enable intelligent exploration of the enormous parameter spaces of materials science—that is, every single sample synthesis and measurement step is selected so as to yield maximal value (scientific insight) such that one can iterate towards a desired material property, or a answer a desired scientific question, as rapidly as possible.

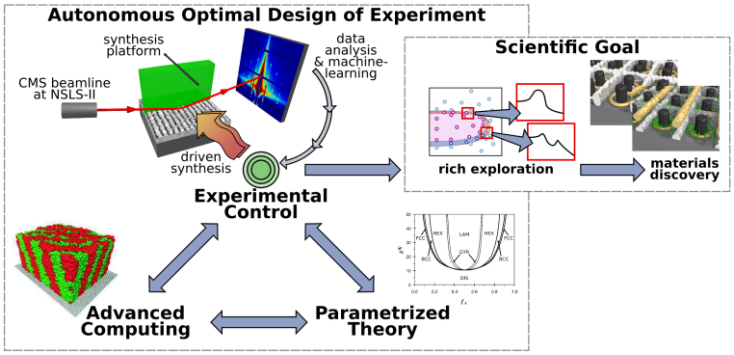
Recent work



Brookhaven National Laboratory has focused on developing these concepts and deploying them in the context of x-ray scattering, which is a powerful and rapid probe of material structure that can even measure materials in-situ (as they are being synthesized or processed). In this work, we have developed analytic[1, 2] and deep learning[3-7] methods for classifying or healing[8, 9] x-ray scattering datasets, and algorithms for optimal decision-making in an experimental context. We have demonstrated how physics-informed deep learning can deliver substantial performance improvements. For instance, we created a multi-channel convolution neural network in which initial data transformations are carefully selected by domain experts to highlight features of interest. For x-ray scattering data, decomposing the raw detector image into a matrix of Fourier-Bessel coefficients efficiently highlights symmetry information in the raw signal.

When these methods are combined and deployed at a synchrotron x-ray scattering beamline, they enable the instrument to autonomously explore material science problems. For instance, the beamline was able to efficiently image the structure of a nanoparticle thin film, where it first measured a low-resolution (coarse) image of the sample makeup, and then selected follow-up experiments so as to achieve higher imaging resolution in areas of interest (especially edges in the image). In another set of experiments, the beamline autonomously measured a large set of sample (using robotics to select samples from a queue) and was able not only to measure these samples in an efficient ordering, but also to suggest what follow-up samples should be synthesized next. Finally, this research program has demonstrated how these methods can be combined with combinatorial sample preparation. For instance, sample libraries can be synthesized by creating continuous gradients (of, e.g., composition); subsequent autonomous study allows these spaces to be mapped efficiently.

Future Needs



A key route towards improvement of the autonomous experimentation paradigm is to enable input of known material physics. This existing understanding both constrains the exploration problem, providing initial estimates of material behavior for guiding exploration, and also provides a framework into which newly-acquired data can be fit. Materials physics can be captured by appropriate simulation tools, such as molecular dynamics or field theories. However, materials models are typically computationally expensive, especially when they are attempting to capture the non-equilibrium aspects of realistic materials, and must thus explicitly simulation material evolution. Thus, a key challenge in the integration of materials models into autonomous workflows is to merge high-performance computing into a real-time experimental context.

Progress will require developments along three key vectors:

1. Software platforms that allow one to seamlessly integrate different computation inputs. In particular, the ability for experiments to select from a menu of decision-making algorithms, and to easily accept input from arbitrary materials modeling code.
2. New models of materials physics must be developed that provide reasonable predictive power while being computationally tractable. Machine-learning approximants can be trained based on the outputs of rigorous models. Ideally infrastructure would be developed to connect models of different fidelity, allowing both rapid input from approximants, as well as intermittent input from expensive models.
3. Infrastructure to enable timely (i.e. *during* experimentation) access to significant computing power, through connecting to existing HPC clusters, or by accessing novel distributed resources. Access must be rapid and elastic, able to handle the inconsistent and 'bursty' nature of experimental data collection, while also scaling favorable to handle the changing complexity of the underlying physics models.

Outlook

Autonomous experimentation has the potential to radical transform scientific study, by liberating human scientists to focus on high-level conceptual understanding, while having scientific instruments automatically handle sample management, processing, and high-speed decision-making. In the future, this paradigm must leverage large computational resources in order to provide real-time inputs from computationally-expensive materials modeling. New cyber infrastructure is critically required to enable timely and cost-effective access to elastic computing resources.

References

1. K. G. Yager and P. W. Majewski, *J. Appl. Crystallogr.*, 2014, **47**, 1855-1865.
2. J. R. Lhermitte, A. Stein, C. Tian, Y. Zhang, L. Wiegart, A. Fluerasu, O. Gang and K. G. Yager, *IUCrJ*, 2017, **4**, 604-613.
3. H. Huang, S. Yoo, K. Kaznatcheev, K. G. Yager, F. Lu, D. Yu, O. Gang, A. Fluerasu and H. Qin, presented in part at the Proceedings of the 29th Annual ACM Symposium on Applied Computing, Gyeongju, Republic of Korea, 2014.
4. M. H. Kiapour, K. Yager, A. C. Berg and T. L. Berg, Applications of Computer Vision (WACV), 2014.
5. B. Wang, Z. Guan, S. Yao, H. Qin, M. H. Nguyen, K. G. Yager and D. Yu, Scientific Data Summit (NYSDS), New York, 2016.
6. B. Wang, K. G. Yager, D. Yu and M. H. Nguyen, Applications of Computer Vision (WACV), 2017.
7. Z. Guan, H. Qin, K. G. Yager, Y. Choo and D. Yu, *British Machine Vision Conference*, 2018, **0828**, 1-10.
8. J. Liu, J. Lhermitte, Y. Tian, Z. Zhang, D. Yu and K. G. Yager, *IUCrJ*, 2017, **4**, 455-465.
9. J. Liu and K. G. Yager, *IUCrJ*, 2018, **5**, 737-752.