CHEMISTRY OF MATERIALS

pubs.acs.org/cm

Editorial

Unveiling the Hidden Influence of Defects via Experiment and Data Science

Cite This: *Chem. Mater.* 2023, 35, 7351−7354

Read Online

ACCESS | Metrics & More | Article Recommendations

The elephant in the room of materials chemistry is that all materials inherently have defects and impurities. Defects can tremendously influence a material's functional properties by deteriorating the performance in some samples, enhancing performance in others, or occasionally contributing to entirely new functionality. A classic example is the presence of intrinsic surface oxygen vacancies (defects) on metal oxide photocatalysts such as TiO_2.[1−3] These surface oxygen vacancies are vital for effectively binding reactant molecules and facilitating charge transfer, boosting activity.[2,4] Similar charge trapping defects also enable applications in luminescent materials.[5] Persistent luminescence phosphors, for example, require the same defects to trap electrons in the excited state, which are eventually slowly released through thermal activation. This defect-related emission process gives rise to the famous glow-in-the-dark property of $SrAl_2O_4:Eu^{2+},Dy^{3+}$ found in emergency signs and children's toys.[6,7] Mitigating these same defects in phosphors enables their application in LED lighting and display applications.[8] Targeting and controlling specific (point) defect formation and concentrations is clearly indispensable for obtaining desired functionalities.

Unfortunately, unless defects are the focus of a research project, most scientists and engineers gloss over their impact for one main reason—they are incredibly difficult to identify, characterize, and understand how they contribute to ensemble-level sample properties. In this editorial, we highlight this "defect challenge", emphasizing that it can only be solved through a cross-cutting research effort supported by computational and experimentalists alike. It will require new techniques in high-throughput computing, data science, and, most importantly, spectroscopy at the single-particle level. Further adopting FAIR (Findable, Accessible, Interoperable, and Reusable) principles[9] and thinking about how large data sets are shared and archived will play a vital role in making use of ultrarich (in information and actual costs) data sets that are currently being under-utilized.

Significant effort is going into establishing new approaches to study defects with the hope of closing this gap. Developments in computational modeling have started to unlock a fundamental understanding of how defects influence a material's properties. Density functional theory (DFT), for example, has provided insight into the defect energy levels and their role in semiconducting and insulating materials like $SrAl_2O_4:Eu^{2+}$ to understand how intrinsic defects enable its famous persistent luminescence.[10] Advanced computational methods have since been developed to more accurately represent the local structure of materials containing point defects, ensuring the lowest energy geometries are modeled.[11] Similar progress has been made in considering multiple defects, defect clustering, and defect dynamics,[12,13] while machine learning force fields in molecular dynamics are gaining popularity for capturing the materials chemistry of larger systems, although the ability for molecular dynamics (MD) to capture defect physics is still limited. Efforts to automate this computational process are also furthering the ability to study the impact of defects on an array of systems, leading to new crystal−chemical connections that can be used to manipulate a material's physical properties. One setback, however, is that these calculations require computationally expensive hybrid functionals to account for charge localization, making it intractable to perform these calculations on a diverse number of systems. Despite this reality, there is still potential for tremendous impact as computational techniques and power continue to improve in their ability to model complex materials with multiple defect types.

Computation has indeed made it possible to capture defect physics more accurately than ever. Yet, current approaches mostly use static supercell models with isolated defects that do not accurately reflect "real" materials. In reality, a single sample containing billions of TiO_2 or $SrAl_2O_4:Eu^{2+},Dy^{3+}$ particles may include individual defects that produce both desirable and undesirable properties, depending on the type and distribution at the single-particle level. This activity heterogeneity effect is also beautifully demonstrated in Pt nanocatalysts. They can exhibit order-of-magnitude variations in single-particle activity,[14] even though the nanocatalysts are nominally identical and were made at the same time in the same reactor. Recent high-resolution 3D electron microscopy measurements have shown that the Pt atom positions vary from particle to particle.[15] The particle-dependent atomic positions cause varying amounts of strain near domain boundaries, dislocation edges, and surface sites, which are strongly linked to the particle-dependent catalytic activity. The devil is in the (synthetic) details—how atoms assemble to form particles influences their individual properties and function. Such observations present a significant challenge for the materials

Published: September 26, 2023

chemistry community: Can we control defect types and locations within particles to unlock extraordinary function?

Researchers wanting to "defect engineer" often do not have the necessary tools, time, and/or massive data and image analysis capabilities to amass defect statistics (types and locations) across order-of-magnitude length scales, from the atomic-level to surface facets, to particle morphologies, to particle−particle interfaces, and eventually the ensemble-level. The community needs to know the atomic-level chemical and physical makeup of particle subpopulations within a sample batch and how the heterogeneous distribution of particles contributes to ensemble-level behavior before they can even begin to engineer desired physical properties. Doing so will finally enable the community to explain performance variations among devices constructed by different laboratories, even though they were constructed from nominally identical materials. Still, we are a long way from this capability.

Progress is being made, though. The first challenge is simply finding the defects in a bulk material. Considerable advances have been made in experimentally visualizing defects in particles using techniques like transmission electron microscopy (TEM). An inevitable limitation of TEM is that even though the atom-scale resolution afforded by today's best microscopes enables resolving structural features on picometer length scales, such a hyper-localized image usually does not represent bulk properties. How often do publications show micrographs of a single feature or a small subset of particles imaged by scanning electron microscopy (SEM) while claiming the data represents the bulk samples? This may be because analysis methods are historically time-consuming and expensive and often focus on highlighting the "best" (defect-free) result. Recent efforts have focused on automating this process; methods include automated electron microscopes, data collection, and data processing, which have expedited micrograph acquisition rates, expanding the analysis beyond tens or hundreds of nanoparticles by orders of magnitude.[16] Regrettably, the ability to image a complete micrometer-sized particle that makes up functional materials today remains impracticable.

Single-particle imaging techniques are emerging to address this gap.[17,18] Sweeping across an entire particle and identifying specific functional "hot spots" within a sample provides information on higher activity regions compared to inactive parts of a particle.[19] These results can be correlated with defect-rich regions in a complete particle, ultimately revealing performance limitations in properties like photocatalysis[20] and Li-ion transport.[21] Similar techniques could be applied to inorganic phosphors to identify portions of the samples with high photoluminescence efficiency relative to the "dead spots" in the sample.

Nevertheless, a drawback of current characterization techniques is that they are limited to a small number of particles, making it challenging to connect defect imaging back to their synthetic origin. However, potential solutions are on the horizon. One promising approach involves merging single-particle imaging with in situ spectroscopy and microscopy characterization tools. Together, these data offer a comprehensive understanding of defect distribution within entire particles. In the future, combining flow synthesis setups (especially in nanoparticle preparation) with these same characterization methods could allow researchers to analyze a large number of particles in real time. This comprehensive analysis would provide insights into particle nucleation, growth, size distributions, shape, crystallinity, and, importantly, defect identification and quantification. Researchers can then establish optimal synthesis−property−defect relationships by varying flow synthesis conditions during data collection. However, it is essential to note that the vast amount of data generated by these experiments would require significant modifications to experimental setups and incorporating data science techniques to manage and interpret the information effectively.

Applying convolutional neural networks and deep learning architectures will be a key advance for researchers to automatically identify and characterize intricate micrograph features, such as particle morphology, crystal defects, or interfacial structures.[22] Supervised learning algorithms can be trained to classify different materials or phases, helping automatically identify and quantify defect locations within complex microstructures. On the other hand, unsupervised learning techniques enable the discovery of previously unknown patterns or relationships, uncovering subtle variations in composition, structure, or morphology. These same techniques can be used to analyze the >terabytes of data likely to be generated when single particle analysis techniques are combined with flow nanoparticle synthesis, *finally* paving the way for a deeper understanding of a material's behavior and structure−property relationships.

Even if these data-driven techniques are adapted to the high-throughput analysis of bulk particle properties, connecting these data back to the synthesis to yield improved materials will remain challenging. Materials synthesis is multifaceted and influenced by temperature, pressure, solvents, and precursors (and their purity), among other variables. Data-driven materials chemistry has been touted to utilize the tremendous amount of data generated by these various imaging techniques as machine learning inputs. Researchers now have the unique ability to use explainable AI or interpret machine learning models to extract information from these high-dimensional algorithms. However, many of these methods only provide a surface-level analysis of the features controlling the model output. There is no guarantee that the information from a SHAP (SHapley Additive exPlanations) or feature importance analysis will provide reliable experimental insight. This is not a limitation for image or defect analysis in materials chemistry but a broader scientific need with significant potential to transform materials synthesis, processing, and the resulting physical properties.

Most importantly, the massive scale of the defect problem requires the production and exchange of large data sets between researchers. Frameworks for sharing computational and, to a lesser extent, experimental data sets are reasonably mature and widely accessed today. Adding information related to the influence of defects would be a reasonable and helpful extension. However, sharing more expansive data sets related to single particle imaging is not currently being done in any organized way. When combined with the metadata, the considerable amount of data generated by these experiments will likely require dedicated hardware and creative ways to share information. Currently, data are commonly shared through repositories like FigShare, but these are limited to 20 GB, and although multiple TB can also be stored, it is costly. Then the question arises: What happens when faculty members decide they no longer want to cover that cost? Government-supported repositories may be a better option, but what happens to the data when funding priorities change?

https://doi.org/10.1021/acs.chemmater.3c01817
Chem. Mater. 2023, 35, 7351−7354

Further, it is necessary to access the algorithms and codes that produced and analyzed the data. Today's convention is to upload codes through repositories like GitHub, which is a great start. However, code maintenance and proper version control, among many other issues, are required to keep everything functioning. These questions must be addressed across the chemistry and materials fields, where data-intensive research is growing dramatically. However, it may be even more vital for areas with tremendously large data sets like single-particle fluorescence microscopy, where the raw images from a single paper exceeded 1 TB and the processed data (in the form of a MATLAB workspace file) is 50 GB.[19] This is just one paper that only studied 37 individual particles and did not include the atomic-level characterization of each particle. The bottom line is that "getting to the bottom" will require massive data generation and storage solutions.

At the end of the day, defects cannot and should not be ignored in materials chemistry. They are pivotal as they significantly influence functional properties with implications spanning all modern technologies. State-of-the-art DFT calculations and imaging techniques, including 3D electron microscopy and high-resolution transmission electron microscopy, have emerged as powerful tools for defect identification. However, these predominantly local analysis methods must readily provide a complete picture of defect distributions within bulk samples or large particles. Fortunately, there is an opportunity to expand beyond traditional materials chemistry by combining flow synthesis with electron microscopy techniques and single-particle imaging methods;[23] real-time observations and analysis of nanoparticle synthesis can be achieved to reveal the optimal synthesis parameters that control defect formation. AI must also be adapted to analyze these data and establish robust correlations between image features and physical properties. Of course, sharing these tremendously large data sets and models following the FAIR principles will be a challenge. Doing this, however, is a worthwhile endeavor to eventually bridge the gap between synthetic processes, experimental data, and data science, although it will require developments in both hardware and software. Overcoming these challenges will yield profound insights into defect distributions, material behavior, and the design of enhanced functional materials.

Justin Sambur orcid.org/0000-0002-8457-4946
Jakoah Brgoch orcid.org/0000-0002-1406-1352

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemmater.3c01817

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

JB would like to thank Dr. Shruti Hariyani for valuable feedback on the manuscript and Prof. David Scanlon and Seán Kavanagh for helpful discussion, as well as the National Science Foundation (DMR-1847701) for supporting this work. JBS acknowledges support from Ensembles of Photosynthetic Nanoreactors (EPN), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science (DE-SC0023431).

REFERENCES

(1) Pan, X.; Yang, M.-Q.; Fu, X.; Zhang, N.; Xu, Y.-J. Defective TiO2 with oxygen vacancies: synthesis, properties and photocatalytic applications. *Nanoscale* **2013**, *5* (9), 3601−3614.
(2) Thompson, T. L.; Yates, J. T. TiO2-based Photocatalysis: Surface Defects, Oxygen and Charge Transfer. *Top. Catal.* **2005**, *35* (3), 197−210.
(3) Pacchioni, G. Oxygen Vacancy: The Invisible Agent on Oxide Surfaces. *ChemPhysChem* **2003**, *4* (10), 1041−1047.
(4) Schaub, R.; Thostrup, P.; Lopez, N.; Lægsgaard, E.; Stensgaard, I.; Nørskov, J. K.; Besenbacher, F. Oxygen Vacancies as Active Sites for Water Dissociation on Rutile TiO2. *Phys. Rev. Lett.* **2001**, *87* (26), No. 266104.
(5) Ho, Y.-C.; Hoque, M. N. F.; Stoneham, E.; Warzywoda, J.; Dallas, T.; Fan, Z. Reduction of Oxygen Vacancy Related Traps in TiO2 and the Impacts on Hybrid Perovskite Solar Cells. *J. Phys. Chem. C* **2017**, *121* (43), 23939−23946.
(6) Xu, J.; Tanabe, S. Persistent luminescence instead of phosphorescence: History, mechanism, and perspective. *J. Lumin.* **2019**, *205*, 581−620.
(7) Huang, K.; Le, N.; Wang, J. S.; Huang, L.; Zeng, L.; Xu, W.-C.; Li, Z.; Li, Y.; Han, G. Designing Next Generation of Persistent Luminescence: Recent Advances in Uniform Persistent Luminescence Nanoparticles. *Adv. Mater.* **2022**, *34* (14), No. 2107962.
(8) Linderälv, C.; Åberg, D.; Erhart, P. Luminescence Quenching via Deep Defect States: A Recombination Pathway via Oxygen Vacancies in Ce-Doped YAG. *Chem. Mater.* **2021**, *33* (1), 73−80.
(9) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; 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. *Scientific Data* **2016**, *3* (1), 160018.
(10) Finley, E.; Mansouri Tehrani, A.; Brgoch, J. Intrinsic Defects Drive Persistent Luminescence in Monoclinic SrAl2O4:Eu2+. *J. Phys. Chem. C* **2018**, *122* (28), 16309−16314.
(11) Mosquera-Lois, I.; Kavanagh, S. R.; Walsh, A.; Scanlon, D. O. Identifying the ground state structures of point defects in solids. *npj Computational Materials* **2023**, *9* (1), 25.
(12) Cen, J.; Zhu, B.; Kavanagh, S. R.; Squires, A. G.; Scanlon, D. O. Cation disorder dominates the defect chemistry of high-voltage LiMn1.5Ni0.5O4 (LMNO) spinel cathodes. *J. Mater. Chem. A* **2023**, *11* (25), 13353−13370.
(13) Krajewska, C. J.; Kavanagh, S. R.; Zhang, L.; Kubicki, D. J.; Dey, K.; Gałkowski, K.; Grey, C. P.; Stranks, S. D.; Walsh, A.; Scanlon, D. O.; Palgrave, R. G. Enhanced visible light absorption in layered Cs3Bi2Br9 through mixed-valence Sn(ii)/Sn(iv) doping. *Chemical Science* **2021**, *12* (44), 14686−14699.
(14) Han, K. S.; Liu, G.; Zhou, X.; Medina, R. E.; Chen, P. How Does a Single Pt Nano-catalyst Behave in Two Different Reactions? A Single-Molecule Study. *Nano Lett.* **2012**, *12* (3), 1253−1259.
(15) Kim, B. H.; Heo, J.; Kim, S.; Reboul, C. F.; Chun, H.; Kang, D.; Bae, H.; Hyun, H.; Lim, J.; Lee, H.; Han, B.; Hyeon, T.; Alivisatos, A. P.; Ercius, P.; Elmlund, H.; Park, J. Critical differences in 3D atomic structure of individual ligand-protected nanocrystals in solution. *Science* **2020**, *368* (6486), 60−67.
(16) Schorb, M.; Haberbosch, I.; Hagen, W. J. H.; Schwab, Y.; Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **2019**, *16* (6), 471−477.
(17) Xu, X.; Valavanis, D.; Ciocci, P.; Confederat, S.; Marcuccio, F.; Lemineur, J.-F.; Actis, P.; Kanoufi, F.; Unwin, P. R. The New Era of

High-Throughput Nanoelectrochemistry. *Anal. Chem.* **2023**, *95* (1), 319−356.
(18) Baker, L. A. Perspective and Prospectus on Single-Entity Electrochemistry. *J. Am. Chem. Soc.* **2018**, *140* (46), 15549−15559.
(19) Sambur, J. B.; Chen, T.-Y.; Choudhary, E.; Chen, G.; Nissen, E. J.; Thomas, E. M.; Zou, N.; Chen, P. Sub-particle reaction and photocurrent mapping to optimize catalyst-modified photoanodes. *Nature* **2016**, *530* (7588), 77−80.
(20) Chen, R.; Ren, Z.; Liang, Y.; Zhang, G.; Dittrich, T.; Liu, R.; Liu, Y.; Zhao, Y.; Pang, S.; An, H.; Ni, C.; Zhou, P.; Han, K.; Fan, F.; Li, C. Spatiotemporal imaging of charge transfer in photocatalyst particles. *Nature* **2022**, *610* (7931), 296−301.
(21) Merryweather, A. J.; Schnedermann, C.; Jacquet, Q.; Grey, C. P.; Rao, A. Operando optical tracking of single-particle ion dynamics in batteries. *Nature* **2021**, *594* (7864), 522−528.
(22) Kalinin, S. V.; Ophus, C.; Voyles, P. M.; Erni, R.; Kepaptsoglou, D.; Grillo, V.; Lupini, A. R.; Oxley, M. P.; Schwenker, E.; Chan, M. K. Y.; Etheridge, J.; Li, X.; Han, G. G. D.; Ziatdinov, M.; Shibata, N.; Pennycook, S. J. Machine learning in scanning transmission electron microscopy. *Nature Reviews Methods Primers* **2022**, *2* (1), 11.
(23) Wang, L.; Schmid, M.; Sambur, J. B. Single nanoparticle photoelectrochemistry: What is next? *J. Chem. Phys.* **2019**, *151* (18), 180901 DOI: 10.1063/1.5124710.

https://doi.org/10.1021/acs.chemmater.3c01817
Chem. Mater. 2023, 35, 7351−7354