DATA-DRIVEN MATERIAL INVESTIGATIONS: UNDERSTANDING FATIGUE BEHAVIOR

Data-Driven Materials Investigations: The Next Frontier in Understanding and Predicting Fatigue Behavior

ASHLEY D. SPEAR^{1,6}, SURYA R. KALIDINDI,² BRYCE MEREDIG,³
ANTONIOS KONTOSOS,⁴ and JEAN-BRIAC LE GRAVEREND⁵

1.—Department of Mechanical Engineering, University of Utah, 1550 Kennecott Building, Salt Lake City, UT 84112, USA. 2.—George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA. 3.—Citrine Informatics, Redwood City, CA, USA. 4.—Mechanical Engineering and Mechanics, Drexel University, Philadelphia, PA, USA. 5.—Aerospace Engineering, Texas A&M University, College Station, TX, USA. 6.—e-mail: ashley.spear@utah.edu

INTRODUCTION

In their seminal work, Hey et al.¹ identify four distinct paradigms that represent the historical evolution of science and technology: (1) empirical science, (2) theoretical-model science, (3) computational science (simulations), and (4) the emerging paradigm of data-driven science. A recent perspective from Agrawal and Choudhary² points out that this sequence of paradigms appears throughout the materials science community specifically and corresponds to advancements over time in the ability to observe, interpret, and represent material behavior. In the first two paradigms, the representation of physical or mechanical behavior is typically of low dimension and is limited to those behaviors with relatively simple governing physics. Prime examples include elasticity and yield in structural metals, where the governing physics and the corresponding salient features of the material structure have been identified to a high level of confidence. More complicated governing physics typically require numerical approaches to predict physical phenomena with reasonable confidence. This spurred the evolution of the third paradigm of computational science. Even with improvements through this third paradigm, the space of the governing physics (including all model forms and parameters needed to connect the material structure to phenomena of interest) that needs to be explored systematically can be quite large. Furthermore, the available experimental data for model validation are rather small in terms of number of

distinct test cases, yet can be rich in terms of data dimensionality for a given test case. Thus, the fourth paradigm of data-driven science is simultaneously motivated and enabled by the relatively recent and unprecedented access to multidimensional and multimodal data derived from sophisticated experiments, simulations, and combinations thereof. New ways to mine, leverage, and manage these data are now required, and addressing this requirement is foundational to the Materials Genome Initiative.³

As with other areas of materials science and engineering, the area of fatigue of materials has closely followed the four paradigms outlined above. The first paradigm was based strictly on empirical approaches and included, for example, the development of the Basquin⁴ and Coffin-Manson^{5,6} formulas to describe, respectively, the stress-life (S-N) or strain-life (ϵ -N) response, as well as the Paris Law^{7,8} to relate crack-growth rate to cyclic stress-intensity factor. Such approaches are limited in their predictive power and provide little insight into the mechanisms governing cyclic deformation, crack nucleation, and crack propagation. The second paradigm was marked by the derivation of theoretical fatigue models, which included, for example, closed-form energy-based models relating crack initiation to persistent slip bands.⁹ Such theoretical models provide more mechanistic insight but remain limited in their range of applicability. The third paradigm has been characterized by the development of computational models to replicate or predict fatigue behavior. Included in the third paradigm are simulations based on dislocation dynamics and crystal plasticity. The third paradigm has provided a much higher-fidelity representation and improved understanding of fatigue phenomena.

Ashley Spear, Antonios Kontsos, and Jean-Briac le Graverend are the guest editors for the Mechanical Behavior of Materials Committee of the TMS Structural Materials Division for the topic Data-driven Material Investigations: Understanding Fatigue Behavior in this issue.

However, the ability to fully predict the evolution of fatigue failure by accounting for microstructure-dependent scatter in the very early stages of fatigue-crack evolution (including incubation, nucleation, and microstructurally small propagation) remains a challenging and unresolved task. This task is further complicated by the amount and type of data that are typically produced by current computational simulations and advanced experimental measurements, which can prove unwieldy to manage, process, and analyze.

This *JOM* special topic focuses specifically on the emerging fourth paradigm as it pertains to the study and prediction of fatigue behavior. With this introductory article and the papers that follow, we hope to provide a platform to highlight both the challenges and opportunities associated with the analysis of large and/or complex data sets and how researchers in the fatigue community are working to address these challenges through data-driven approaches. In the following subsections, we present more details on the emerging paradigm of data-driven science and present cases where such science has facilitated materials discovery, enhanced or accelerated characterization of materials, and led to new mechanistic understanding. We then describe opportunities for leveraging data-driven approaches in the fatigue community, and we close by offering recommendations moving forward.

PARADIGM OF DATA-DRIVEN MATERIALS SCIENCE

In general, data-driven approaches aim to identify objectively (relying largely on the available data) the embedded correlations among selected inputs and outputs needed to study or model a given phenomenon. Data-driven approaches are particularly advantageous when a full, physics-based understanding of relevant phenomena is lacking, and also when gathering new experimental data is particularly slow and costly.¹⁰ The relatively recent fourth paradigm of data-driven materials science has been or is being realized in various areas of materials science. Efforts in this paradigm are mainly focused on extracting high-value information from all available materials data (generated by either experiments or computations) and expressing it as high-value linkages among material processes, structures, and properties. This is especially well suited for practical materials-design explorations. The central impediment in this effort arises from the lack of a rigorous mathematical framework for quantifying the material structure, whose salient features span multiple material length scales (from the atomistic to the macroscale). The very large number of parameters needed to fully capture all details of the hierarchical material structure make the structure representation inherently high dimensional. However, from a practical viewpoint, it is essential to be able to identify high-value, low-

dimensional representations of the material structure that can be employed reliably to drive the material innovation efforts leading to enhanced properties. One type of data-driven method that has already made inroads across many subfields of materials research is that of machine learning (ML).^{11,12} Below, we briefly discuss examples of well-established use cases for ML that have been fruitful in areas of materials science.

Materials Discovery

ML-based models can aid materials discovery by predicting materials properties of interest. The models may be trained directly on experimental data^{13,14} or act as very fast surrogates for more expensive physics-based simulation such as density functional theory¹⁵. ML has successfully guided the experimental discovery of novel Heusler alloys,¹⁶ Ni-based superalloys,¹⁷ and shape-memory alloys,¹⁸ among many other application areas.

Data-driven experimental planning, sometimes called sequential learning or active learning, has emerged as a key application for ML in materials science. In such an approach, we cast materials discovery as a fundamentally iterative process and address the question of optimally planning a series of experiments to deliver high-performing materials. ML-driven experiments tend to identify promising materials much more efficiently than a naive search.¹⁹ Several implementations of ML-optimized experimental design have emerged, including COMBO,²⁰ FUELS,¹⁹ and work from Xue et al.¹⁸

Enhanced or Accelerated Characterization

Given the abundance of spectroscopic and image-based materials characterization techniques, researchers in materials informatics have found it fruitful to apply state-of-the-art computer vision techniques²¹ to materials problems. DeCost and Holm²² used convolutional neural networks (CNNs) to predict properties of materials samples based on their microstructures. Ziatdinov et al.²³ automatically identified defects in atomic-scale microscopy data with deep learning. Xu and LeBeau²⁴ employed CNNs to dramatically accelerate the process of analyzing electron beam diffraction patterns. Across these applications, the primary benefits of applying ML to characterization data include (1) extracting more scientific insight from each experiment and/or (2) processing experimental data much more efficiently, with less human time-intensive input.

New Mechanistic Understanding

The materials informatics community has begun to recognize the scientific limitations of black-box ML models and increasingly is exploring questions of model interpretability. In an ideal scenario, ML would help scientists generate improved hypotheses and would assist in the identification of underlying

mechanisms of materials phenomena. Recent work has addressed physical interpretability of ML in the context of micrograph classification,²⁵ graph representations of crystal structure,²⁶ and molecular structures of promising organic solar materials.²⁷ Learning physics with the assistance of data-driven methods is a rapidly emerging area of research that is likely to substantially increase interest in these methods as it will reduce the dissonance between the traditional scientific method—which seeks greater understanding—and the perceived black-box nature of ML.

A ROLE FOR DATA-DRIVEN SCIENCE IN THE FATIGUE COMMUNITY

Indeed, the fatigue community has already made progress through this fourth paradigm,^{28–31} and many opportunities exist to continue to support or leverage data-driven approaches in the pursuit of better understanding and prediction of fatigue behavior. For example, implementation of image-processing algorithms are needed to mine and parameterize—automatically and robustly—the full-field data derived from state-of-the-art experiments or simulations. Opportunities also exist to develop new methods of integrating rich experimental data sets with high-fidelity modeling to enable correlative studies or prediction of fatigue behavior. Similarly, data-driven algorithms are needed to link length and time scales, to correlate multimodal data sets, to systematically identify patterns, and ultimately to predict fatigue behavior. Such algorithms might be applications of ML, which could either enable discovery of fatigue mechanisms or serve as surrogate models to represent the physical mechanisms of fatigue. In the context of fatigue models, data-driven approaches can play an important role in addressing several problems. Specific examples might include (1) rank-ordering the combinations of materials parameters (related to chemistry, microstructure, and processing-history) that exhibit strong correlations to the fatigue life and (2) rank-ordering alternate specifications of the governing physics (including model forms and associated model parameters) based on how they are supported by the available experimental observations. Problems of this type are indeed best answered by data-driven approaches as they are capable of accounting rigorously for the uncertainty associated with the data sets (these can include both experimental and modeling data sets). In other words, these methods fully recognize that one cannot provide deterministic answers to the exemplary problems identified above within the constraints of the available limited data, but one can provide the answers in a much more practically useful probabilistic framework. Therefore, the potential benefits and value of employing the data-driven approaches in addressing challenging problems in

fatigue are clearly self-evident. However, much work still needs to be done in learning how to best employ the existing toolsets with any needed modifications to facilitate their efficient application.

The following papers presented in this issue of *JOM* provide cutting-edge examples of how researchers are currently pursuing data-driven approaches in fatigue applications. To download any of the papers, follow the url (<https://link.springer.com/journal/11837/70/7>) to the table of contents page for the July 2018 issue (vol. 70, no. 7).

- “A data-analytics approach for discovering unique microstructural configurations susceptible to fatigue” by S. K. Jha, R. A. Brockman, R. M. Hoffman, V. Sinha, A. L. Pilchak, W. J. Porter, D. J. Buchanan, J. M. Larsen, and R. John.
- “Data-driven mechanistic modeling of microstructural influence on the high cycle fatigue life of nickel titanium” by O. L. Kafka, C. Yu, M. Shakoar, Z. Liu, G. J. Wagner, and W. K. Liu.
- “Data-driven correlation analysis between observed 3D fatigue-crack path and computed fields from high-fidelity, crystal-plasticity, finite-element simulations” by K. D. Pierson, J. D. Hochhalter, and A. D. Spear.
- “Data-science analysis of the macro-scale features governing the corrosion to crack transition in AA7050-T7451” by N. E. C. Co, D. E. Brown, and J. T. Burns.
- “Visualization and quantitative analysis of crack-tip plastic zone in pure nickel” by R. D. Kelton, J. Fathi, E. Meletis, and H. Huang.
- “Fatigue damage assessment leveraging nondestructive evaluation data” by K. Mazur, B. J. Wisner, and A. Kotsos.

RECOMMENDATIONS TO FOSTER DATA-DRIVEN FATIGUE MODELING

In closing, the authors would like to offer three concrete recommendations that would help to facilitate the adoption and accountability of data-driven methodologies among the fatigue community. First, the community should create several publicly accessible benchmark data sets, which could be used to test and directly compare models derived from data-driven approaches. An example of such a published materials data set is the UltraHigh Carbon Steel Micrograph DataBase (UHCSDB).³² Second, the community should organize blind-prediction competitions to assess data-driven fatigue models, analogous to the Sandia Fracture Challenge^{33,34} or the long-running organic crystal structure prediction blind tests.³⁵ Finally, the community should adopt data platforms such as Citrination,³⁶ Materials Commons,³⁷ and Materials Data Facility,³⁸ as these systems greatly facilitate data and model sharing, reproducibility of results, and reuse of code.

ACKNOWLEDGEMENTS

A.D.S. would like to acknowledge support from the Air Force Office of Scientific Research under award number FA9550-15-1-0172. S.R.K. and A.D.S. would also like to acknowledge support from the National Science Foundation under grant nos. 1333083 and 1629660, respectively.

REFERENCES

1. T. Hey, S. Tansley, and K.M. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, vol. 1 (Redmond: Microsoft Research, 2009).
2. A. Agrawal and A. Choudhary, *Apl Mater.* 4(5), 053208 (2016).
3. J.P. Holdren, *Materials Genome Initiative for Global Competitiveness* (Washington: National Science and Technology Council OSTP, 2011).
4. O. Basquin, *Proceedings of ASTM*, vol. 10 (1910), pp. 625–630.
5. L.F. Coffin Jr., *Trans. ASME* 76, 931 (1954).
6. S.S. Manson, Behavior of materials under conditions of thermal stress, National Advisory Commission on Aeronautics: report 1170 (1954).
7. P. Paris, M. Gomez, and W. Anderson, *Trend Eng.* 13(1), 9 (1961).
8. P. Paris and F. Erdogan, *J. Basic Eng.* 85(4), 528 (1963).
9. K. Tanaka and T. Mura, *J. Appl. Mech.* 48(1), 97 (1981).
10. M.L. Hutchinson, E. Antono, B.M. Gibbons, S. Paradiso, J. Ling, and B. Meredig, Overcoming data scarcity with transfer learning (2017). [arXiv:1711.05099](#).
11. T. Mueller, A.G. Kusne, and R. Ramprasad, *Rev. Comput. Chem.* 29, 186 (2015).
12. J. Hill, G. Mulholland, K. Persson, R. Seshadri, C. Wolverton, and B. Meredig, *MRS Bull.* 41(5), 399 (2016).
13. V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, and I. Takeuchi, Machine learning modeling of superconducting critical temperature (2017). [arXiv:1709.02727](#).
14. M.W. Gaultois, A.O. Oliynyk, A. Mar, T.D. Sparks, G.J. Mulholland, and B. Meredig, *APL Mater.* 4(5), 053213 (2016).
15. A. Khorshidi and A.A. Peterson, *Comput. Phys. Commun.* 207, 310 (2016).
16. A.O. Oliynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, and A. Mar, *Chem. Mater.* 28(20), 7324 (2016).
17. B. Conduit, N. Jones, H. Stone, and G. Conduit, *Mater. Des.* 131, 358 (2017).
18. D. Xue, R. Yuan, Y. Zhou, D. Xue, T. Lookman, G. Zhang, X. Ding, and J. Sun, *Sci. Rep.* 6, 28244 (2016).
19. J. Ling, M. Hutchinson, E. Antono, S. Paradiso, and B. Meredig, *Integr. Mater. Manuf. Innov.* 6, 207 (2017).
20. T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, and K. Tsuda, *Mater. Discov.* 4, 18 (2016).
21. A. Krizhevsky, I. Sutskever, and G.E. Hinton, *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.
22. B. L. DeCost and E. A. Holm, *Comput. Mater. Sci.* 110, 126 (2015).
23. M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R.R. Unocic, R. Vasudevan, S. Jesse, and S.V. Kalinin, *ACS Nano* 11(12), 12742 (2017).
24. W. Xu and J.M. LeBeau, A deep convolutional neural network to analyze position averaged convergent beam electron diffraction patterns (2017). [arXiv:1708.00855](#).
25. J. Ling, M. Hutchinson, E. Antono, B. DeCost, E.A. Holm, and B. Meredig, Building data-driven models with microstructural images: generalization and interpretability (2017). [arXiv:1711.00404](#).
26. T. Xie and J.C. Grossman, Crystal graph convolutional neural networks for accurate and interpretable prediction of material properties (2017). [arXiv:1710.10324](#).
27. S.A. Lopez, B. Sanchez-Lengeling, J. de Goes Soares, and A. Aspuru-Guzik, *Joule* 1, 857 (2017).
28. A. Agrawal, P.D. Deshpande, A. Cecen, G.P. Basavarsu, A.N. Choudhary, and S.R. Kalidindi, *Integr. Mater. Manuf. Innov.* 3, 8 (2014).
29. M.W. Priddy, N.H. Paulson, S.R. Kalidindi, and D.L. McDowell, *Int. J. Fatigue* 104, 231 (2017).
30. N. Paulson, *Structure-Property Linkages for Polycrystalline Materials using Materials Knowledge Systems*. PhD Thesis, Georgia Institute of Technology (2017).
31. A. Rovinelli, M.D. Sangid, H. Proudhon, Y. Guilhem, R.A. Lebensohn, and W. Ludwig, *J. Mech. Phys. Solids* 115, 208 (2018).
32. B.L. DeCost, M.D. Hecht, T. Francis, B.A. Webler, Y.N. Picard, and E.A. Holm, *Integr. Mater. Manuf. Innov.* 6(2), 197 (2017).
33. B.L. Boyce, S.L.B. Kramer, H.E. Fang, T.E. Cordova, M.K. Neilsen, K. Dion, A.K. Kaczmarowski, E. Karasz, L. Xue, A.J. Gross, A. Ghahremaninezhad, K. Ravi-Chandar, S.-P. Lin, S.-W. Chi, J.S. Chen, E. Yreux, M. Rüter, D. Qian, Z. Zhou, S. Bhamare, D.T. O'Connor, S. Tang, K.I. Elkhodary, J. Zhao, J.D. Hochhalter, A.R. Cerrone, A.R. Ingraffea, P.A. Wawrzyniek, B.J. Carter, J.M. Emery, M.G. Veilleux, P. Yang, Y. Gan, X. Zhang, Z. Chen, E. Madenci, B. Kilic, T. Zhang, E. Fang, P. Liu, J. Lua, K. Nahshon, M. Miraglia, J. Cruce, R. DeFrese, E.T. Moyer, S. Brinckmann, L. Quinkert, K. Pack, M. Luo, and T. Wierzbicki, *Int. J. Fract.* 186(1–2), 5 (2014).
34. B.L. Boyce, S.L.B. Kramer, T.R. Bosiljjevac, E. Corona, J.A. Moore, K. Elkhodary, C.H.M. Simha, B.W. Williams, A.R. Cerrone, A. Nonn, J.D. Hochhalter, G.F. Bomarito, J.E. Warner, B.J. Carter, D.H. Warner, A.R. Ingraffea, T. Zhang, X. Fang, J. Lua, V. Chiaruttini, M. Mazière, S. Feld-Payet, V.A. Yastrebov, J. Besson, J.-L. Chaboche, J. Lian, Y. Di, B. Wu, D. Novokshanov, N. Vajragupta, P. Kucharczyk, V. Brinnel, B. Döbereiner, S. Münstermann, M.K. Neilsen, K. Dion, K.N. Karlson, J.W. Foulk III, A.A. Brown, M.G. Veilleux, J.L. Bignell, S.E. Sanborn, C.A. Jones, P.D. Mattie, K. Pack, T. Wierzbicki, S.-W. Chi, S.-P. Lin, A. Mahdavi, J. Predan, J. Zdravcevec, A.J. Gross, K. Ravi-Chandar, and L. Xue, *Int. J. Fract.* 198(1–2), 5 (2016).
35. A.M. Reilly, R.I. Cooper, C.S. Adjiman, S. Bhattacharya, A.D. Boese, J.G. Brandenburg, P.J. Bygrave, R. Bylsma, J.E. Campbell, R. Car et al., *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* 72(4), 439 (2016).
36. J. OMara, B. Meredig, and K. Michel, *JOM* 68(8), 2031 (2016).
37. B. Puchala, G. Tarcea, E.A. Marquis, M. Hedstrom, H. Jagadish, and J.E. Allison, *JOM* 68(8), 2035 (2016).
38. B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke, and I. Foster, *JOM* 68(8), 2045 (2016).