Materials LAB

Perspective

Perspectives in the New Era of Materials Intelligent Design

Published as part of the Virtual Special Issue "Beihang University at 70"

Ruifeng Zhang[1,2*], Tengfei Xu[1,2], Bonan Yao[1,2], and Zhaorui Liu[1,2]

[1] *School of Materials Science and Engineering, Beihang University, Beijing 100191, China*

[2] *Center for Integrated Computational Engineering (International Research Institute for Multidisciplinary Science) and Key Laboratory of High-Temperature Structural Materials & Coatings Technology (Ministry of Industry and Information Technology), Beihang University, Beijing 100191, China*

* Corresponding author, E-mail: zrf@buaa.edu.cn

Abstract

The launching integrated computational materials engineering (ICME) and materials genome engineering (MGE) has led the transformation of empirical and theoretical design paradigm into the rational computational one that further provides the basis for the data-driven design paradigm by integrating the high-throughput techniques in experiments and computations, the big data science with general principles, the informatics with knowledge discovery based on data mining and machine learning, and ultimately enabling the possibility of materials intelligence design (MID) via artificial intelligence. In this perspective article, we highlight the intelligent solution to acquire the processing-structure-property-performance relationship of multilevel-structured materials by emphasizing modularization, automation, standardization, integration and intelligence, following the hierarchical relationship of data, information, knowledge and wisdom, which is essentially different from the past empirical, theoretical and computational paradigms. The new era of MID is expected to fundamentally reform the material innovation mode through an integrated infrastructure guided by novel concepts that is radically distinguished from the way of thinking and doing in the past, providing a perspective scientific vision and direction for future materials design.

Key words: Materials genome engineering; integrated computational materials engineering; materials intelligent design; high-throughput techniques; machine learning

Citation: Ruifeng Zhang, Tengfei Xu, Bonan Yao, Zhaorui Liu. Perspectives in the New Era of Materials Intelligent Design. *Materials Lab* 2022, 1, 220017. DOI: 10.54227/mlab.20220017

• Main text

With the rapid development of modern materials modeling, computation and simulation, it has gradually become a reality to integrate computational material design into the whole product system design[1] and to promote the generation of massive data in materials, and thus the empirical and theoretical design paradigm is transforming to the rational computational design one, and further to intensively develop data-driven design paradigm according to the concepts of ICME and MGE.[2] To realize the unified contemporary design of materials and products, ICME is to provide an integration of computational materials science and product system by combining the material information and knowledge obtained by computational simulation with product analysis and processing at different scales.[3] Because of the inherent feature of hierarchy and heterogeneity of materials from the extreme range of spatial and temporal scales, the physics of each scale is dominated by a set of fundamental microvariables, which characterize the collective behavior of those at smaller scales. As a result, it requires to bridge multiscale methods across scales, e.g., the message passing approach and the embedding scheme,[4] which depends strongly on the capabilities of materials computations & simulations. To reach the targeted materials intelligent design (MID), MGE aims to put forward a solution for the material design through the generation of big data via high-throughput techniques, the physical and mathematic solutions to obtain materials information and knowledge, and an intelligent expert/decision system for prediction and design, fundamentally triggering the transformation of scientific and technological innovation mode. In the realization of this proposal, the high-throughput computations & simulations, experimental preparation & characterization are essential for the establishment of materials database, which in turn builds a foundation for data analysis & mining (e.g., machine learning, ML), thus providing physical or mathematic models through information extraction and knowledge acquisition, and finally promoting the construction of an integrated MID system.

Figure 1 presents a unified research and design strategy of future material design, which brings a new era of MID by integrating the high-throughput techniques, data science, informatics with knowledge discovery and artificial intelligence (AI) with the launching of ICME and MGE.[5–7] As shown in the bottom-left of Figure 1, to meet the requirements of MID, the multiscale high-throughput computations & simulations (e.g., first principles calculations based on density functional theory)[8] and high throughput experimental preparations &

Received 24 March 2022; Accepted 25 April 2022; Published online

characterizations (e.g., combinatorial chip technology[9]), provide surely a basis for the data-driven design paradigm. Such techniques are tied with the idea that automate the whole process of calculations and simulations, as well as the syntheses and characterizations by emphasizing the integration, modularization, automation, standardization, and parallel & concurrent processing with high-efficiency. So far various high throughput computation & simulation tools are freely available in public, e.g., Materials Project,[10] Automatic Flow for materials discovery (AFLOW),[11] SPaMD Studio[12] and meanwhile, some representative data repositories have substantially boosted, including the computational ones such as AFLOWLIB and Open Quantum Materials Database (OQMD),[13] as well as those experimental ones, e.g., the High-Throughput Experimental Materials Database (HTEM DB) of inorganic materials. [14]

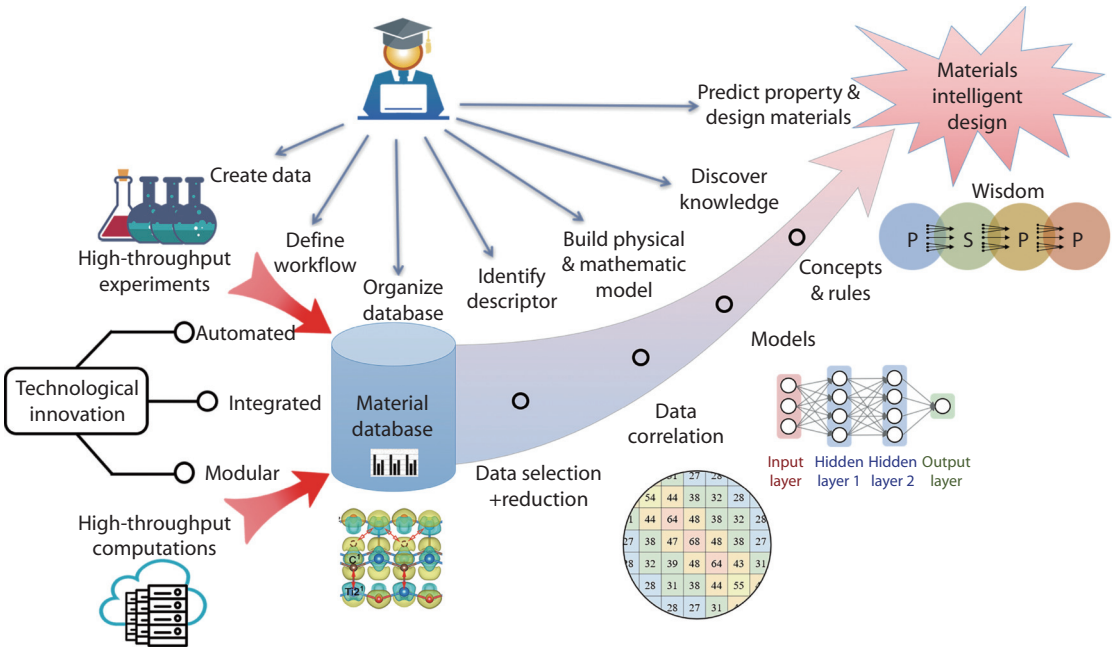

Fig. 1 The unified research and design strategy of future material design, which brings a new era of MID by integrating the high-throughput techniques, data science, informatics with knowledge discovery and the artificial intelligence to acquire the processing-structure-property-performance (PSPP) relationship.[5–7]

With massive raw data in ready, as depicted in the middle of Figure 1, the materials informatics provides a general framework for the data analysis & mining of useful information, which may further help the knowledge-discovery to build physical or chemical models, concepts or rules through exploring the data correlation (i.e., linear or non-linear equation with physical significance), e.g., material "fingerprint" (also known as "descriptor", i.e., the critical structural and/or physical parameter that depict a certain property of the materials[15]), or to build various mathematic models based on machine learning approaches, e.g., artificial neural networks (ANN).[16] The core role of machine learning should be recognized in data mining that is based on mathematics, statistics, informatics, visualization and other disciplines, since it provides the unreachable capability for physical modelling and prediction with consideration of the multiscale complexity of materials behaviors and multilevel approximations in computations & simulations. In this aspect, several successful ML examples include the ML potentials for molecular dynamics simulations,[17] the ML models of structure-property relationships for property predictions,[18] etc.

By integrating the data science and the informatics with knowledge discovery, the MID is feasibly implemented through building AI system, as presented in the top-right of Figure 1. The AI is generally termed as the computer expert/decision system for prediction and inverse design that applies the physical models, principles and rules, or mathematical models obtained by machine learning yet cannot be clearly expressed. Nevertheless, MID provides a new scientific vision for future materials innovations, thus is necessary to combine various fields including the big data science with high throughput computational & experimental methods, the materials informatics with knowledge discovery via data-mining and machine learning, and the artificial intelligence system through expert/decision system, etc., to promote insightful interdisciplinary exchanges of idea, methods, and results, in turn to speed up future materials discovery by transforming material research into an integrated intelligent platform for material prediction and design.

In short, ICME and MGE provide a basis of novel concepts and directions of R&D in the field of material science and engineering in recent years by effectively integrating high-throughput computations & simulations, high-throughput experimental preparations & characterizations, the storage and organization of big data with general principles, the information extraction and the knowledge discovery based on data mining and machine learning, as well as an artificial intelligence system for prediction and design. Following the hier-

DOI: 10.54227/mlab.20220017

archical relationship of data, information, knowledge and wisdom (DIKW),[6] the MID paradigm allows one to acquire the PSPP relationship by emphasizing modularization, automation, standardization, integration and intelligence, and apply it as the basis for the rational materials design in a faster, cheaper, more accurate and intelligent way. In this new scientific vision, the material innovation is fundamentally different from the way of thinking and doing in the past empirical, theoretical and computational paradigms, and it needs to be supported by a whole new infrastructure that at least includes the integrated frameworks for high-throughput experimentation and high-throughput computation, the robust facilities for big data storage and organization, the informatics tools with the capabilities for various data analysis and mining, and an intelligent system for prediction and design. Last but not least, big challenges are clearly faced in MID, including the veracity, relevance, completeness, standardization of big data, the automation, modulation, and integration of high-throughput methodologies, the efficiency and generalization of algorithms to build physical and mathematic models, and the AI system, and a broad collaboration with sufficient sharing of resources, data, and algorithms, etc.

Acknowledgments

The authors acknowledge the National Key Research and Development Program of China (Grants No. 2017YFB0702100), "111 Project" (No. B17002), National Thousand Young Talents Program of China, and Fundamental Research Funds for the Central Universities.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Ruifeng Zhang: Conceptualization, Writing-Original Draft, Writing-Review & Editing, Supervision, Funding acquisition. Tengfei Xu: Writing-Review & Editing. Bonan Yao: Writing-Review & Editing. Zhaorui Liu: Writing-Review & Editing. All authors had approved the final version of the manuscript.

REFERENCES

1. D. L. McDowell and G. B. Olson, *Sci. Model. and Simul.*, 2008, 15, 207
2. National Science and Technology Council Committee on Technology Subcommittee on the Materials Genome Initiative, Materials Genome Initiative Strategic Plan, USA, June 2014.
3. National Research Council, *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security*, The National Academies Press, USA, 2008.
4. R. Lesar, *Introduction to Computational Materials Science: Fundamentals to Applications*, Cambridge University Press, UK, 2013.
5. T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research Press, USA, 2009.
6. S. R. Kalidindi, *Int. Mater. Rev.*, 2015, 60, 150
7. A. Agrawal and A. Choudhary, *APL Mater.*, 2016, 4, 053208
8. G. Ceder and K. Persson, *Sci. Am.*, 2013, 309, 36
9. H. Wang, X. D. Xiang and L. T. Zhang, *Sci. Technol. Rev.*, 2018, 36, 15
10. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholla, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, 1, 011002
11. S. Curtarolo, W. Setyawan, G. L. Hart, M. Jahnatek, R. V. Chepulskii, R. H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy and M. J. Mehl, *Comput. Mater. Sci.*, 2012, 58, 218
12. Z. R. Liu, B. N. Yao, R. F. Zhang, *Comput. Mater. Sci.*,
13. S. Kirklin, J. E. Saal, B. Meredig, A. Tompson, J. W. Doak, M. Aykol, S. Ruhl and C. Wolverton, *npj Comput. Mater.*, 2015, 1, 1
14. A. Zakutayev, N. Wunder, M. Schwarting, J. D. Perkins, R. White, K. Munch, W. Tumas and C. Phillips, *Sci. Data*, 2018, 5, 1
15. L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, *Phys. Rev. Lett.*, 2015, 114, 105503
16. D. J. Livingstone, Artificial neural networks: Methods and Applications, Humana Press, USA, 2008.
17. J. Behler, *Chem. Rev.*, 2021, 121, 10037
18. G. L. W. Hart, T. Mueller, C. Toher, and S. Curtarolo, *Nat. Rev. Mater.*, 2021, 6, 730

Biography

Ruifeng Zhang is a full professor of the School of Materials Science and Engineering at Beihang University, China. He obtained Ph.D. in Materials Science from Tsinghua University in 2005 with honors of "Excellent PhD Graduate" and "Excellent PhD Thesis" of the university, and received Alexander von Humboldt Fellowship (Germany) and Los Alamos Director's Postdoctoral Fellowship (USA). He has published over 150 papers in international journals such as Chem. Rev., Phys. Rep., PNAS, PRL, Adv. Mater., Acta. Mater. with >4000 citations (h-index=43). Several software in computational materials science are released in public, e.g. MiedCalc; AELAS, ADAIS, PNADIS; SPaMD, EAPOT, AACSD, AADIS.