How would AI shape CALculation of PHAse Diagrams?

Anonymous Author(s)
Affiliation
Address
email

Abstract

While the CALPHAD which stands for CALculation of PHAse Diagrams is one of the most successful data-driven paradigms and has already made huge impact in materials science especially physical metallurgy during past 50 years, AI-based approaches recently are getting more and more attention in many different application scenarios within computational materials science. In this attention track perspective, we would like to share the considerations about the possibility of integrating the current trendy AI-based methods into CALPHAD mainly for obtaining high-quality data and the thermodynamic model development.

1 Introduction

CALPHAD, a methodology introduced in 1970 by Larry Kaufman is perhaps one of the earliest methods integrating physical modeling and learning from data to digitize the thermodynamics of materials with phase diagram which is the graphical representations of the distributions of the phases under the different external conditions [1, 2, 3]. The CALPHAD workflow can be roughly divided into the following sections: data capture, construction of the thermodynamic model, optimization of the model through updating the undetermined parameters,database generation and application to many cases such as: phase stability prediction [4, 5], phase-field modeling [6, 7], precipitation simulation [8, 9], etc. However, current CALPHAD has several challenges: first, lack of the high-quality data; second, the commonly used thermodynamic model is simple but not very robust; Third, due to the possible inconsistency caused by the multiple sources of data, it's hard to automatically determine the optimized modeling but with lots of artifacts.

While huge progress in many scientific disciplinaries is made by AI especially the neural-network (NN) based machine learning (ML) recently, it's especially effective for the multiscale phenomena with curse of dimensionality [10] such as protein folding [11], fluid mechanics [12] and quantum many-body simulations [13]. Though the thermodynamics is typically not considered as a multiscale problem, its complexity still comes from the huge amount of the degree of freedoms. Here we would like to mainly discuss the possibility of whether AI-led methods would assist CALPHAD to deal with its current challenges including lack of the data and developing more robust thermodynamic model.

2 AI for data side of CALPHAD

One of the key challenges of the current CALPHAD is the lack of the necessary data, which includes thermochemical data and phase boundary data. However, experimental data acquisition is always a tedious and time-consuming process. As a result, the first-principles calculations based on density functional theory (DFT) have been widely equipped to supplement the experimental data and reach the acceptable accuracy for thermochemical data [14, 15]. Nevertheless, required phase boundary data is still hard to be obtained directly from the first-principles calculations. The main challenge to

Submitted to NeurIPS 2021 AI for Science Workshop, Attention Track.

calculate the phase boundaries is it requires high computational accuracy: an error of 1 meV/atom in Gibbs energy of a condensed matter phase may cause an error around 10 K in transition temperature [16, 17]. The error of the current state-of-the-art (SOTA) phase boundary estimation fully based on DFT without ML is around 100K [16, 18] and this is far more away from replacing the experimental measurement to provide the predictions for technological purpose. One way to deal with this accuracy issue is to go beyond DFT to attempt more accurate first principles methods [19, 17, 20]. However, it is usually accompanied with huge computational cost and the wider application is limited.

With the help of current NN-based ML, it is possible to tackle the problem by more efficiently constructing DFT functionals approaching the chemical accuracy and several recent works have already focused on this path [21, 22]. However, we need to consider solid solution [23], liquid phase with ordering [24], while in realistic experiments, anharmonicity [25] and defects [26] always exist in materials, only DFT-level electronic structure calculation perhaps cannot handle all this and need the larger scale simulation. That's exactly the case we need the multiscale modeling to reach the required accuracy with ML as a tool to across the scale gap [10]. One recent progress is the satisfied phase diagrams produced by DeePMD which is a kind of molecular dynamics simulation with the ML trained ab initio potentials [27, 28]. It can be hypothesized that if we connect all the SOTA computational methods at the different scales with the help of ML, it would have the chance to produce accurate enough high-throughput phase boundaries data all based on first principles calculations and resolve the lack of the experimental data for CALPHAD.

3 Model side of CALPHAD: a specialized ML model?

Only data is not enough, CALPHAD still needs robust, easy-to-use thermodynamic models to depict the stability of the different phases in the target system. Here the thermodynamic model means to develop a description for the free energy of a phase as a function of some degrees of freedom related to our interest. We consider the thermodynamic model within CALPHAD itself could be seen as a kind of specialized ML model. It is trained with limited amount of the collected data to identify the correct phase stability while CALPHAD modeling here could be seen as a supervised learning task. It is also worth to mention that the cluster variation method (CVM) [29, 30], which is considered as a potential thermodynamics model for CALPHAD but limited on its complexity, is introduced to gain the insight to develop the generalized belief propagation to study the probabilistic inference [31, 32, 33]. This implies the strong relationship between the thermodynamic model and machine learning [34]. Recently CALPHAD community expects to impose more physics into CALPHAD and replace the empirical model to improve its robust [35]. However, the current popular NN based ML models usually requires huge amount of data and that's exactly one of the challenges of CALPHAD, lack of the high-quality data. Even with the larger dataset, the NN-based model is still suspicious to apply as it generally not considers the physics constraint here is the thermodynamics law. Some recent work is taking use of non NN-based methods within CALPHAD community but only on some specific cases such as the uncertainty quantifications [36, 37]. The rapidly developing few-shot learning [38] may be proper for this scenario, but the related attempt is still lacking.

4 Conclusion

Based on all we have discussed, we consider current CALPHAD would definitely benefit from AI through obtaining high-quality data. There are still some other connections between CALPHAD and AI-led methods we haven't discussed: automatic data collection with text-mining and database construction [39, 40, 41], high-throughput thermodynamics data generation [42, 43], constructing data ecosystem for data recycle [44], AI-guided high-throughput experimental autonomy for CALPHAD [45], etc. However, just like what we mentioned they are still mainly about the data side of CALPHAD. As we consider the thermodynamic model within CALPHAD can be treated as a specialized ML model, the development of more robust thermodynamic model is necessary but still a gap. This is another interesting topic to explore and would perhaps benefit the both sides.

References

[1] Larry Kaufman and Harold Bernstein. Computer calculation of phase diagrams. with special reference to refractory metals. 1970.

[2] Bo Sundman, HL Lukas, and SG Fries. *Computational thermodynamics: the Calphad method*. Cambridge university press New York, 2007.

[3] Zi-Kui Liu and Yi Wang. *Computational thermodynamics of materials*. Cambridge University Press, 2016.

[4] KJ Hubbard and DG Schlom. Thermodynamic stability of binary oxides in contact with silicon. *Journal of Materials Research*, 11(11):2757–2776, 1996.

[5] Zi-Kui Liu, DG Schlom, Qi Li, and XX Xi. Thermodynamics of the mg–b system: Implications for the deposition of mgb 2 thin films. *Applied Physics Letters*, 78(23):3678–3680, 2001.

[6] JZ Zhu, ZK Liu, Venu Vaithyanathan, and LQ Chen. Linking phase-field model to calphad: application to precipitate shape evolution in ni-base alloys. *Scripta Materialia*, 46(5):401–406, 2002.

[7] Kaisheng Wu, YA Chang, and Yunzhi Wang. Simulating interdiffusion microstructures in ni–al–cr diffusion couples: a phase field approach coupled with calphad database. *Scripta materialia*, 50(8):1145–1150, 2004.

[8] Xiangyu Xia, Weihua Sun, Alan A Luo, and Donald S Stone. Precipitation evolution and hardening in mgsmznzr alloys. *Acta Materialia*, 111:335–347, 2016.

[9] A. Perron, C. Toffolon-Masclet, X. Ledoux, F. Buy, T. Guilbert, S. Urvoy, S. Bosonnet, B. Marini, F. Cortial, G. Texier, C. Harder, V. Vignal, Ph. Petit, J. Farré, and E. Suzon. Understanding sigma-phase precipitation in a stabilized austenitic stainless steel (316nb) through complementary calphad-based and experimental investigations. *Acta Materialia*, 79:16–29, 2014.

[10] E Weinan, Jiequn Han, and Linfeng Zhang. Integrating machine learning with physics-based modeling. *arXiv preprint arXiv:2006.02619*, 2020.

[11] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[12] Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477–508, 2020.

[13] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[14] Y Wang, S Curtarolo, Chao Jiang, R Arroyave, T Wang, G Ceder, L-Q Chen, and Z-K Liu. Ab initio lattice stability in comparison with calphad lattice stability. *Calphad*, 28(1):79–90, 2004.

[15] Zi-Kui Liu. First-principles calculations and calphad modeling of thermodynamics. *Journal of phase equilibria and diffusion*, 30(5):517–534, 2009.

[16] Li-Fang Zhu, Blazej Grabowski, and Jörg Neugebauer. Efficient approach to compute melting properties fully from ab initio with application to cu. *Physical Review B*, 96(22):224202, 2017.

[17] Florian Dorner, Zoran Sukurma, Christoph Dellago, and Georg Kresse. Melting si: beyond density functional theory. *Physical review letters*, 121(19):195701, 2018.

[18] Li-Fang Zhu, Fritz Körmann, Andrei V Ruban, Jörg Neugebauer, and Blazej Grabowski. Performance of the standard exchange-correlation functionals in predicting melting properties fully from first principles: Application to al and magnetic ni. *Physical Review B*, 101(14):144108, 2020.

[19] Ester Sola and Dario Alfe. Melting of iron under earth's core conditions from diffusion monte carlo free energy calculations. *Physical Review Letters*, 103(7):078501, 2009.

[20] Max Rang and Georg Kresse. First-principles study of the melting temperature of mgo. *Physical Review B*, 99(18):184103, 2019.

[21] Mihail Bogojeski, Leslie Vogt-Maranto, Mark E Tuckerman, Klaus-Robert Müller, and Kieron Burke. Quantum chemical accuracy from density functional approximations via machine learning. *Nature communications*, 11(1):1–11, 2020.

[22] Yixiao Chen, Linfeng Zhang, Han Wang, and Weinan E. Deepks: A comprehensive data-driven approach toward chemically accurate density functional theory. *Journal of Chemical Theory and Computation*, 17(1):170–181, 2020.

[23] Axel van de Walle and Gerbrand Ceder. Automating first-principles phase diagram calculations. *Journal of Phase Equilibria*, 23(4):348–359, 2002.

[24] Chandler A Becker, John Ågren, Marcello Baricco, Qing Chen, Sergei A Decterov, Ursula R Kattner, John H Perepezko, Gernot R Pottlacher, and Malin Selleby. Thermodynamic modelling of liquids: Calphad approaches and contributions from statistical physics. *physica status solidi (b)*, 251(1):33–52, 2014.

[25] Brent Fultz. Vibrational thermodynamics of materials. *Progress in Materials Science*, 55(4):247–352, 2010.

[26] Jutta Rogal, Sergiy V Divinski, Mike W Finnis, Albert Glensk, Joerg Neugebauer, John H Perepezko, Sergej Schuwalow, Marcel HF Sluiter, and Bo Sundman. Perspectives on point defect thermodynamics. *physica status solidi (b)*, 251(1):97–129, 2014.

[27] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and E Weinan. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters*, 120(14):143001, 2018.

[28] Haiyang Niu, Luigi Bonati, Pablo M Piaggi, and Michele Parrinello. Ab initio phase diagram and nucleation of gallium. *Nature communications*, 11(1):1–9, 2020.

[29] Ryoichi Kikuchi. A theory of cooperative phenomena. *Phys. Rev.*, 81:988–1003, Mar 1951.

[30] R. Kikuchi and K. Masuda-Jindo. Cluster variation method in the computational materials science. *Calphad*, 26(1):33–54, 2002.

[31] Jonathan S Yedidia, William T Freeman, Yair Weiss, et al. Generalized belief propagation. In *NIPS*, volume 13, pages 689–695, 2000.

[32] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on information theory*, 51(7):2282–2312, 2005.

[33] Alessandro Pelizzola. Cluster variation method in statistical physics and probabilistic graphical models. *Journal of Physics A: Mathematical and General*, 38(33):R309, 2005.

[34] Tom Conte, Erik DeBenedictis, Natesh Ganesh, Todd Hylton, John Paul Strachan, R Stanley Williams, Alexander Alemi, Lee Altenberg, Gavin Crooks, James Crutchfield, et al. Thermodynamic computing. *arXiv preprint arXiv:1911.01968*, 2019.

[35] WA Oates, H Wenzl, and T Mohri. On putting more physics into calphad solution models. *Calphad*, 20(1):37–45, 1996.

[36] Richard A Otis and Zi-Kui Liu. High-throughput thermodynamic modeling and uncertainty quantification for icme. *JOM*, 69(5):886–892, 2017.

[37] Noah H Paulson, Brandon J Bocklund, Richard A Otis, Zi-Kui Liu, and Marius Stan. Quantified uncertainty in thermodynamic modeling for materials design. *Acta Materialia*, 174:9–15, 2019.

[38] Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3), June 2020.

[39] Edward Kim, Kevin Huang, Adam Saunders, Andrew McCallum, Gerbrand Ceder, and Elsa Olivetti. Materials synthesis insights from scientific literature via text extraction and machine learning. *Chemistry of Materials*, 29(21):9436–9444, 2017.

[40] Olga Kononova, Tanjin He, Haoyan Huo, Amalie Trewartha, Elsa A Olivetti, and Gerbrand Ceder. Opportunities and challenges of text mining in aterials research. *Iscience*, 24(3), 2021.

[41] Axel van de Walle, Chiraag Nataraj, and Zi-Kui Liu. The thermodynamic database database. *Calphad*, 61:173–178, 2018.

[42] Logan Ward, Ruoqian Liu, Amar Krishna, Vinay I Hegde, Ankit Agrawal, Alok Choudhary, and Chris Wolverton. Including crystal structure attributes in machine learning models of formation energies via voronoi tessellations. *Physical Review B*, 96(2):024104, 2017.

[43] Adam M Krajewski, Jonathan W Siegel, Jinchao Xu, and Zi-Kui Liu. Extensible structure-informed prediction of formation energy with improved accuracy and usability employing neural networks. *arXiv preprint arXiv:2008.13654*, 2020.

[44] Zi-Kui Liu. Ocean of data: integrating first-principles calculations and calphad modeling with machine learning. *Journal of Phase Equilibria and Diffusion*, 39(5):635–649, 2018.

[45] Wendao Li, Longfei Li, Stoichko Antonov, Changdong Wei, Ji-Cheng Zhao, and Qiang Feng. High-throughput exploration of alloying effects on the microstructural stability and properties of multi-component coni-base superalloys. *Journal of Alloys and Compounds*, page 160618, 2021.