PEROVSKITE PHOTOVOLTAICS

A dataquake for solar cells

The advancement of perovskite photovoltaics has led to a large increase in the volume of published data, which is not always easy to find or reuse. Now, researchers have consistently formatted parameters related to fabrication and performance of over 42,000 solar cells and made them available for analysis in an open-access database.

Marina S. Leite

The fast development of perovskite photovoltaics has led to an enormous number of scientific publications containing a colossal amount of data regarding material fabrication, device layout and processing, and testing conditions. While a large amount of data is paramount to further advance the field, finding the relevant information and then filtering and using it for planning experiments is challenging and time consuming. The ability to access all this information in a single platform, with the possibility of instantly visualizing trends in the data, would enable its analysis at an unparalleled speed and possibly reveal otherwise hidden correlations.

Writing in *Nature Energy*, T. Jesper Jacobsson and Eva Unger from Helmholtz-Zentrum Berlin and Uppsala University, together with 92 other colleagues across the world, now offer such an ability through the Perovskite Database Project[1]. This massive undertaking consists of an open-source online database that contains nearly all information regarding halide perovskites for solar cells published to date in the peer-reviewed literature along with a tool for visualizing and analysing data.

The researchers manually collected published data and information in a consistent format, cataloguing more than 42,000 perovskite solar cells, with almost 100 parameters per device and encompassing over 400 perovskite compositions. Consistent formatting is key to the usability of data, ensuring that the information is shared in a findable, accessible, interoperable, and reusable manner, following what are called the FAIR Data Principles[2]. The database is also intended to be a dynamic and constantly evolving tool. The researchers encourage others in the field to further contribute to the database as they obtain new results. The manuscript contains instructions on how to upload additional information in a consistent format.

The database and its tools can be used to identify the effects of a certain parameter on device stability and performance as well as hidden trends between published data, all of which would help researchers in the field to plan their future experiments. Unger and colleagues provide some examples. For instance, they show the possibility to track the evolution of device power conversion efficiency as a function of time, starting with the very first report of a halide perovskite solar cell. This type of information is very useful to down-select the best performing chemical compositions or carrier transport layers, ultimately guiding device development. The correlations in performance can be analysed as a function of the perovskite composition (for example, $CsPbI_3$ perovskites only) or the type of substrate (flexible or rigid). The database also provides a user-friendly way to visualize specific information about material synthesis and its relationship with device performance. As an example, Unger and colleagues analyse the effects of distinct electron transport layers on the open-circuit voltage and how power conversion efficiency is affected by the presence or absence of a mesoporous TiO_2 layer.

By considering areas where either data is limited or information is not correctly reported, the work also pinpoints potential roadblocks for the field. One of the identified challenges is the urgent need for subcells with optimal optical bandgap for tandem

Credit: GelatoPlus/Getty Images

integration[3]. Another current hold-up is related to scalability[4]. The vast majority of reported efficiencies are measured on devices with an active area of around 0.1 cm^2 and, to a lesser extent, of 1.0 cm^2. However, because of the very limited data on devices with an active area larger than 5.0 cm^2, no conclusive trend could be extrapolated regarding the average performance at large scale. Finally, a major roadblock for the commercialization of perovskite photovoltaics is their relatively short long-term stability. Here, the database can assist researchers in the field in advancing their understanding of the physical and chemical processes that lead to device degradation. To acquire meaningful datasets, it is key that scientists in the field adhere to the International Summit on Organic Photovoltaic Stability (ISOS) protocols, which ensure consistency in the way device stability is assessed and reported[5,6].

The database could also be a key tool for implementing artificial intelligence (AI) methods in the perovskite field. These methods could truly accelerate our knowledge about this class of materials, including determining stable chemical compositions, optimizing material synthesis parameters, understanding the role of environmental stressors in material and device degradation, identifying ideal photovoltaic operating and resting conditions, and defining standard aging test conditions[7]. By applying AI algorithms, the high-dimensional perovskite parameter space could be analysed in a timely manner, unravelling and quantifying patterns and trends that would remain hidden otherwise. In the future, AI could be used to forecast device performance, which would represent a major leap forward in identifying reliable device designs and operating conditions that maximize their lifetime.

This paper provides an inspiring example by taking the first big step towards what should become common practice in the field: making published data easily reusable by the scientific community. Nevertheless, a major challenge for accurate representation of the data lies in accessing and analysing the results of failed experiments[8], which are often not reported in peer-reviewed publications. This information could provide valuable information to train AI algorithms more accurately by discerning between optimal and non-optimal experimental conditions. Together with data from successful experiments, they would make datasets more comprehensive, boosting the potential of AI techniques to guide follow-up experiments. One option to make these unsuccessful findings available is to create additional, free-access databases in a similarly formatted manner for its easy analysis.

The dataquake presented in this work could enable an extraordinary advancement of our understanding of halide perovskites. Furthermore, Unger and colleagues made the code underlying the database and its tools openly accessible so that it can serve as a blueprint for other database projects in the renewable energy space. In light of the growth in the volume of research data across many fields, the Perovskite Database Project may lead the way in managing and exploiting large datasets. ❐

Marina S. Leite ✉

Department of Materials Science and Engineering, University of California, Davis, CA, USA.

✉*e-mail: mleite@ucdavis.edu*

Published online: 13 December 2021
https://doi.org/10.1038/s41560-021-00954-y

References

1. Jacobsson, T. J. et al. *Nat. Energy* https://doi.org/10.1038/s41560-021-00941-3 (2021).
2. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
3. Duong, T. et al. *Adv. Energy Mater.* **7**, 1700228 (2017).
4. Li, Z. et al. *Nat. Rev. Mater.* **3**, 18017 (2018).
5. Khenkin, M. V. et al. *Nat. Energy* **5**, 35–49 (2020).
6. Christians, J. A., Habisreutinger, S. N., Berry, J. J. & Luther, J. M. *ACS Energy Lett.* **3**, 2136 (2018).
7. Srivastava, M., Howard, J. M., Gong, T., Dias, M. R. S. & Leite, M. S. *J. Phys. Chem. Lett.* **12**, 7866 (2021).
8. Raccuglia, P. et al. *Nature* **533**, 73–76 (2016).

Competing interests

The author declares no competing interests.