Prediction ability and importance of descriptors of single-element physical sputtering yields based on sparse modeling

[1]*Hiori Kino, [2]Kazumasa Ikuse, [3]Hieu Chi Dam, & [2]Satoshi Hamaguchi

[1] National Institute for Materials Science, Japan

[2] Center for Atomic and Molecular Technologies, Osaka University, Japan

[3] Japan Advanced Institute of Science and Technology, Japan

**kino.hiori@nims.go.jp*

Due to the rapid development of semiconductor device technologies, industrial demands for more precise and efficient microfabrication processes are increasing at a rapid pace. Plasma processes offer a possible solution to meet such demands. Physical sputtering is one of basic plasma processes but its detailed microscopic mechanisms are not known completely and predicting the physical sputtering yield of a material even of a single element by single-species ion impact is not straightforward.

In this study, we introduce our attempt to predict physical sputtering yields (i.e., etch rates) of various materials by energetic ion impact with a data-driven method. For the sake of simplicity, we focus on a material of single element and ions of single species. Our data is taken from an empirical formula of sputtering yields derived from experimental results for a large variety of substrate elements, incident ion species, and incident ion energies, with the sputtering yield being the target variable. Our analysis is based on the sparse modeling with the exhaustive search method, which selects combinations of descriptors. One of the important points of the exhaustive search method is that it provides many possible combinations of descriptors that offers almost the same prediction accuracy. This may be considered as an advantage. We can choose any such combination of descriptors via an indicator diagram, depending on our purpose, for example, accuracy or computational cost. However, it can be also considered as a disadvantage. There is no clear criterion that defines the best combination of descriptors. Therefore we cannot rate the importance of descriptors or descriptor groups even if we can choose an appropriate combination of descriptors that satisfies our purpose. The subgroup relevance method, where the importance of each descriptor or descriptor group is related directly to the accuracy of the regression, has been invented to overcome this difficulty.* With this method, the prediction ability is shown to depend on the substrate element, incident ion species, and incident ion energy, and the importance of descriptors or descriptor groups gives an insight into the difference between a data-driven method and a theory-driven method.

* Hieu Chi Dam, Viet Cuong Nguyen, Tien Lam Pham, Anh Tuan Nguyen, Kiyoyuki Terakura, Takashi Miyake and Hiori Kino, Journal of the Physical Society of Japan, 87, 113801 (2018). DOI:10.7566/JPSJ.87.113801