Contents lists available at ScienceDirect

Data in Brief

Data Article

Data sets and trained neural networks for Cu migration barriers

Jyri Kimari [a],[*], Ville Jansson [a], Simon Vigonski [a],[b], Ekaterina Baibuz [a], Roberto Domingos [c], Vahur Zadin [b], Flyura Djurabekova [a]

[a] Helsinki Institute of Physics and Department of Physics, University of Helsinki, P.O. Box 43 (Pietari Kalmin katu 2), FI-00014, Finland
[b] Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia
[c] Instituto Politécnico de Nova Friburgo – Universidade do Estado do Rio de Janeiro, Rua Sao Francisco Xavier, 524, 20550–900 Rio de Janeiro, RJ, Brazil

A R T I C L E   I N F O

A B S T R A C T

Kinetic Monte Carlo (KMC) is an efficient method for studying diffusion. A limiting factor to the accuracy of KMC is the number of different migration events allowed in the simulation. Each event requires its own migration energy barrier. The calculation of these barriers may be unfeasibly expensive. In this article we present a data set of migration barriers on for nearest-neighbour jumps on the Cu surfaces, calculated with the nudged elastic band (NEB) method and the tethering force approach. We used the data to train artificial neural networks (ANN) in order to predict the migration barriers for arbitrary nearest-neighbour Cu jumps. The trained ANNs are also included in the article. The data is hosted by the CSC IDA storage service.

© 2020 Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

|	
Subject	Materials Science
Specific subject area	Kinetic Monte Carlo simulations on the surface, assisted by machine learning
Type of data	Table
How data were acquired	Nudged elastic band method with semi-empirical potentials, Cascade2 and iRPROP algorithm for obtaining the neural network weights
Data format	Raw
Parameters for data collection	Hopping events on Cu surfaces of the orientations {100}, {110} and {111} were considered.
Description of data collection	Data sets of migration barriers for surface migration jumps on three different surface orientations were generated with the nudged elastic band method. Artificial neural networks were trained for migration barrier regression on these three surfaces using the Fast Artificial Neural Networks (FANN) library. A classifier network was trained with the same library to assign local atomic environments to one of these three surfaces.
Data source location	University of Helsinki, Finland
Data accessibility	Repository name: CSC IDA storage
	Data identification number:
	urn:nbn:fi:att:a4120c3b-9535–405d-b768–4a972bce0b1b
	Direct URL to data:
	http://urn.fi/urn:nbn:fi:att:a4120c3b-9535–405d-b768–4a972bce0b1b
Related research article	J. Kimari, V. Jansson, S. Vigonski, E. Baibuz, R. Domingos, V. Zadin, F. Djurabekova, Application of artificial neural networks for rigid lattice kinetic Monte Carlo studies of Cu surface diffusion, Computational Materials Science 183 (2020): 109,789. https://doi.org/10.1016/j.commatsci.2020.109789

Value of the Data

- Migration energy barriers are a necessary component of any kinetic Monte Carlo simulations of diffusion. In surface diffusion, the barriers are determined by the configuration of neighbouring atoms – the local atomic environment. Depending on the desired accuracy of the model, these environments may be too many to have their barriers calculated individually. Hence, performing regression with machine learning is an attractive approach.
- The barrier data sets and the trained neural networks can be of interest to computational physicists. The data sets can also be used computer scientists developing any supervised machine learning method.
- The neural networks are ready to be used in computer simulations of Cu surface diffusion. In development of machine learning methods, the trained neural networks can be taken as a benchmark for assessing the performance of new models.

1. Data Description

Two kinds of data are presented in this article. Firstly, there are the tables of migration energy barriers for various nearest-neighbour jumps on Cu surface. Secondly, there are the presentations of the artificial neural networks trained on the barrier data described above.

The barrier data is in three files: Cu_20160906_100.26d, Cu_20160906_110.26d and Cu_20160906_111.26d. The basis for separating the data in these files is the specific surface on which each migration barrier was calculated – the {100}, the {110} or the {111} surface. These sets were treated as separate data sets when training the neural networks in this article, but in principle the data can be combined if desired for some purpose. The encoding of the local atomic environment (LAE) is equivalent in each subset (see Fig. 1). The format is such that on each line there is the 26-dimensional (26D) encoding as a string of 0 s and 1 s – 0 denotes a vacancy, 1 denotes a Cu atom, with an ordering corresponding to the indices shown in Fig. 1– followed by the associated migration energy barrier in eV.

The neural networks are saved in the format defined by the Fast Artificial Neural Networks (FANN) library [1]. They can be loaded with the fann_create_from_file() function. The loading functionality has been implemented in the Kimocs KMC program [2]; the network files are ready
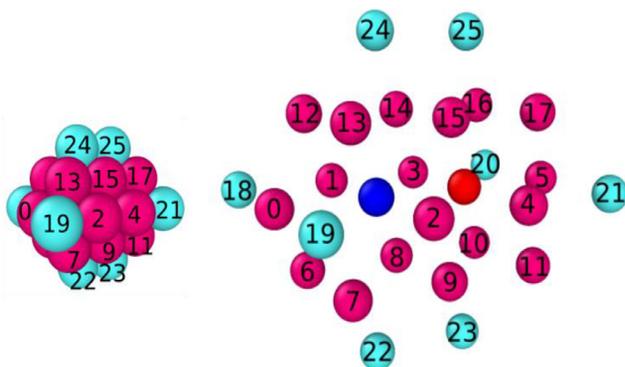
Fig. 1. Encoding of the 26D LAE of a nearest-neighbour jump. The left-hand side displays the lattice sites as touching spheres, the right-hand side shows the sites spread out for clarity. The atom, coloured blue in the right-hand side image, jumps to the lattice site, coloured red. The purple sites are the first-nearest neighbours and the light blue sites are the second-nearest neighbours of the initial or the final position of the jumping atom. The indexing of the 26 neighbour sites from 0 to 25 is shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Symmetry operations associated with each plane of symmetry. Horizontal and vertical refer to the directions as shown in Fig. 1. The planes parallel to the jump direction preserve the jump direction, leaving the barrier unchanged; mirroring perpendicular to the jump direction changes the barrier.

Symmetry plane	Parallel to the jump direction, "horizontal"	Parallel to the jump direction, "vertical"	Perpendicular to the jump direction
Element pairs to be swapped	6 – 12, 7 – 13, 8 – 14, 9 – 15, 10 – 16, 11 – 17, 22 – 24, 23 – 25	1 – 2, 3 – 4, 6 – 7, 8 – 9, 10 – 11, 12 – 13, 14 – 15, 16 – 17, 18 – 19, 20 – 21	0 – 5, 1 – 3, 2 – 4, 6 – 10, 7 – 11, 12 – 16, 13 – 17, 18 – 20, 19 – 21, 22 – 23, 24 – 25

to be used with Kimocs. In total, sixteen different artificial neural networks are presented. Fifteen are regressors that have a 26-dimensional input (the encoded LAE) and a 1-dimensional output (the migration energy in eV). These networks are saved in files Cu_20180209_<100, 110 or 111>_<1, 2, 3, 4 or 5>.net. Five networks correspond to each of the surfaces ({100}, {110}, and {111}). The sixteenth network is a classifier that has 26D input and 3D output – each of the three outputs correspond to the (unnormalised) likelihood of the input LAE to belong in each of the surfaces. This network is saved in Cu_20171023_classifier.net. See Fig. 2 for a flowchart describing how to use these sixteen neural networks in a KMC simulation. This is the procedure followed in the KMC simulations of [3]. Obtaining a migration energy barrier from the combination of artificial neural networks during KMC. The transformation of the LAE refers to the procedure where the representative case from the symmetrically equivalent family is chosen. See text and Table 1 for details.

2. Experimental Design, Materials, and Methods

The migration barriers have been calculated with the nudged elastic band (NEB) method [4,5] implemented in the LAMMPS molecular dynamics program [6]. The potential energy function was described by the molecular dynamics/Monte Carlo corrected effective medium (MD/MC-CEM) potential by Stave et al. [7]. In addition to the interatomic potential, a tethering force was applied to each atom during NEB; see [8] for details. The force constant of the tethering spring was 2.0 eV/Å2. The simulation box size was approximately 45 Å by 45 Å by 25 Å, but the
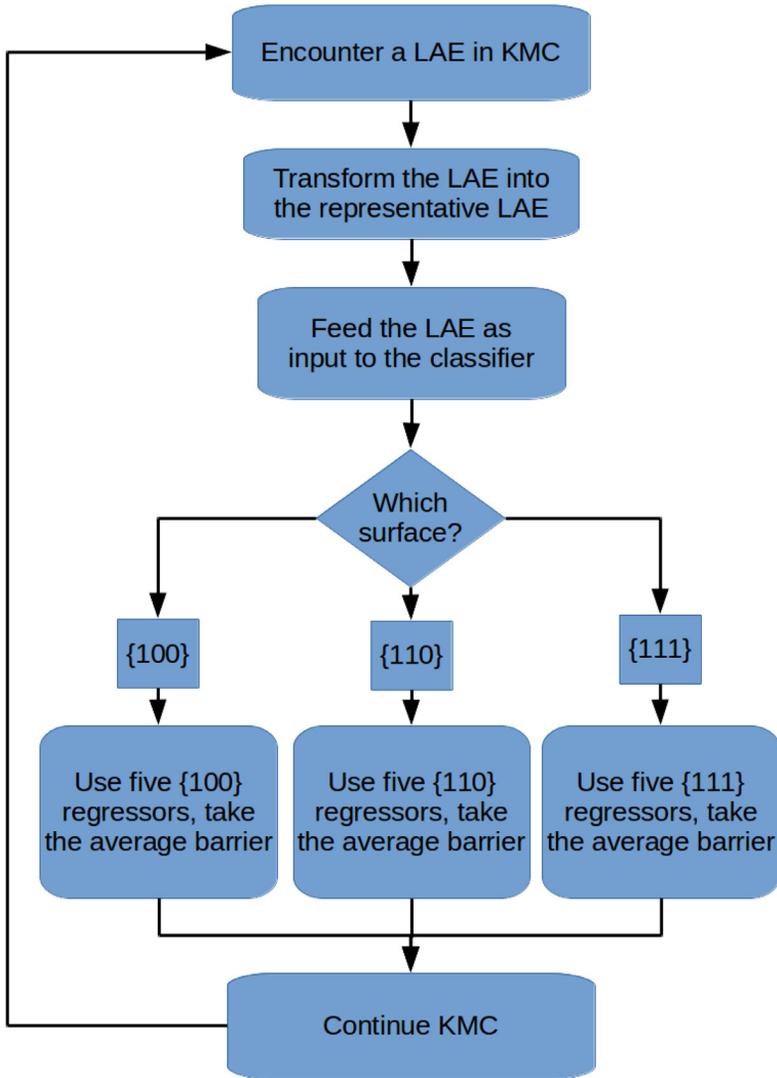
Fig. 2. Obtaining a migration energy barrier from the combination of artificial neural networks during KMC. The transformation of the LAE refers to the procedure where the representative case from the symmetrically equivalent family is chosen. See text and Table 1 for details.

exact dimensions varied according to the surface orientation used in each calculation. The box beyond the LAE was filled with Cu, so that the LAE was embedded on an open surface. The two bottom atomic layers of the box were fixed. Eleven replicas were used in the NEB calculations. For further details, see [3].

The regressor ANNs are cascade multilayer perceptron (MLP) networks with 30–70 hidden nodes. The activation functions of the hidden nodes are sigmoids, and the activation function at the output node is linear. Training was done for until desired accuracy of 0.08 meV (in the case of the {100} and the {110} data sets) or network size of 70 nodes (in the case of the {111} data set} was reached. The classifier ANN is a fully connected MLP network with 30 hidden nodes in one hidden layer, trained with the iRPROP algorithm.

Prior to training, the barrier data sets were modified by removing the symmetrically redundant LAEs and barriers. The LAE has two planes of symmetry parallel to the jump direction (see Fig. 1), so up to four different LAE encodings correspond to the same physically equivalent environment. There is a third plane of symmetry that is perpendicular to the jump direction and will thus change the barrier. See Table 1 for the exact specification of the symmetry operations. The symmetry reduction for the regressors was done so that from each group of up to four equivalent LAEs the encoding with the smallest value was selected as the representative to remain in the data set. The value is defined as the numeric value of the 26-dimensional vector when its elements are concatenated to a single number. For training the classifier, the representative was taken from the entire family of up to eight LAEs; including the reverse processes. This way, the forward and the reverse process barriers will always be taken from the same regressor, regardless of the quality of the classifier.

The symmetries have to be taken into account when using the network. This is marked in the flowchart of Fig. 2 as "Transform the LAE into the representative LAE". This means that whenever a LAE is encountered in KMC, the eight mirror images of that LAE are first generated, and then the one with the smallest value is selected and given as input to the classifier instead of the original LAE. After classification, the smallest image with the same jump direction (not considering the reverse processes in this stage) is given as input for the appropriate regressors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgements

References

[1] S. Nissen, Implementation of a Fast Artificial Neural Network Library (FANN), Department of Computer Science University of Copenhagen (DIKU), 2003 Tech. rep.URL http://fann.sf.net .

[2] V. Jansson, E. Baibuz, F. Djurabekova, Long-term stability of Cu surface nanotips, Nanotechnology 27 (26) (2016) 265708.

[3] J. Kimari, V. Jansson, S. Vigonski, E. Baibuz, R. Domingos, V. Zadin, F. Djurabekova, Application of artificial neural networks for rigid lattice kinetic Monte Carlo studies of Cu surface diffusion, Comput. Mater. Sci. 183 (2020) 109789, doi:10.1016/j.commatsci.2020.109789.

[4] G. Mills, H. Jónsson, Quantum and thermal effects in H2 dissociative adsorption: evaluation of free energy barriers in multidimensional quantum systems, Phys. Rev. Lett. 72 (7) (1994) 1124.

[5] G. Mills, H. Jónsson, G.K. Schenter, Reversible work transition state theory: application to dissociative adsorption of hydrogen, Surf. Sci. 324 (2) (1995) 305–337.

[6] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. 117 (1) (1995) 1–19.

[7] M.S. Stave, D.E. Sanders, T.J. Raeker, A.E. DePristo, Corrected effective medium method. V. simplifications for molecular dynamics and Monte Carlo simulations, J. Chem. Phys. 93 (6) (1990) 4413–4426.

[8] E. Baibuz, S. Vigonski, J. Lahtinen, J. Zhao, V. Jansson, V. Zadin, F. Djurabekova, Migration barriers for surface diffusion on a rigid lattice: challenges and solutions, Comput. Mater. Sci. 146 (2018) 287–302, doi:10.1016/j.commatsci.2017.12.054.