Electronic Supplementary Material (ESI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2023

Machine-learning prediction on formation of atomic gold wires by mechanically controlled break junction

Aishwaryo Ghosh*[a], Biswajit Pabi*[a], Atindra Nath Pal#[a], Tanusri Saha-Dasgupta⊕[a]

September 21, 2023

Supporting Information

S1 Conductance measurement protocol

Before starting any data set (of 1000 traces), to get rid of any memory effect from the local structure, we first manually break and form junction few times using the coarse knob and then choose the triangular piezo pulse with sufficiently large amplitude such that the conductance could reach up to > 200-300 G_0 before each trace. The representative piezo ramp and the corresponding data from a data set (50 mV bias and 3 nm/s stretching speed) are shown in Figure S1. This procedure leads to spontaneously formed fresh junctions at each stretching event. We can see that the upper limit of measurable conductance is 38 G_0 for an applied bias of 50 mV with an amplifier (current to voltage) gain of 10^5 V/A. This is because our 24 bit DAQ card (PCI 4461, National Instrument) can measure a maximum voltage of +/- 10V. However, in reality it is a few 100^{ths} of G_0, ensuring the fact that we indeed rupture any local structure before taking any new trace. Futhermore, Figure S1a also shows the three consecutive sets of conductance traces as a function of piezo bias. Unique evolution of the junctions is better visualized from the zoomed view as shown in Figure S1b. While trace 2 reveals the

*The authors contributed equally.
[a] *S.N. Bose National Centre for Basic Sciences, JD Block, Sector III, Salt Lake, Kolkata 700106, India,*
#: *E-mail: atin@bose.res.in* ⊕ *E-mail: t.sahadasgupta@gmail.com*

chain-like structure, trace 1 and trace 3 are limited to a single atomic junction. A completely different structure near 2.0 G_0 further verify the absence of any possible structural memory effect.

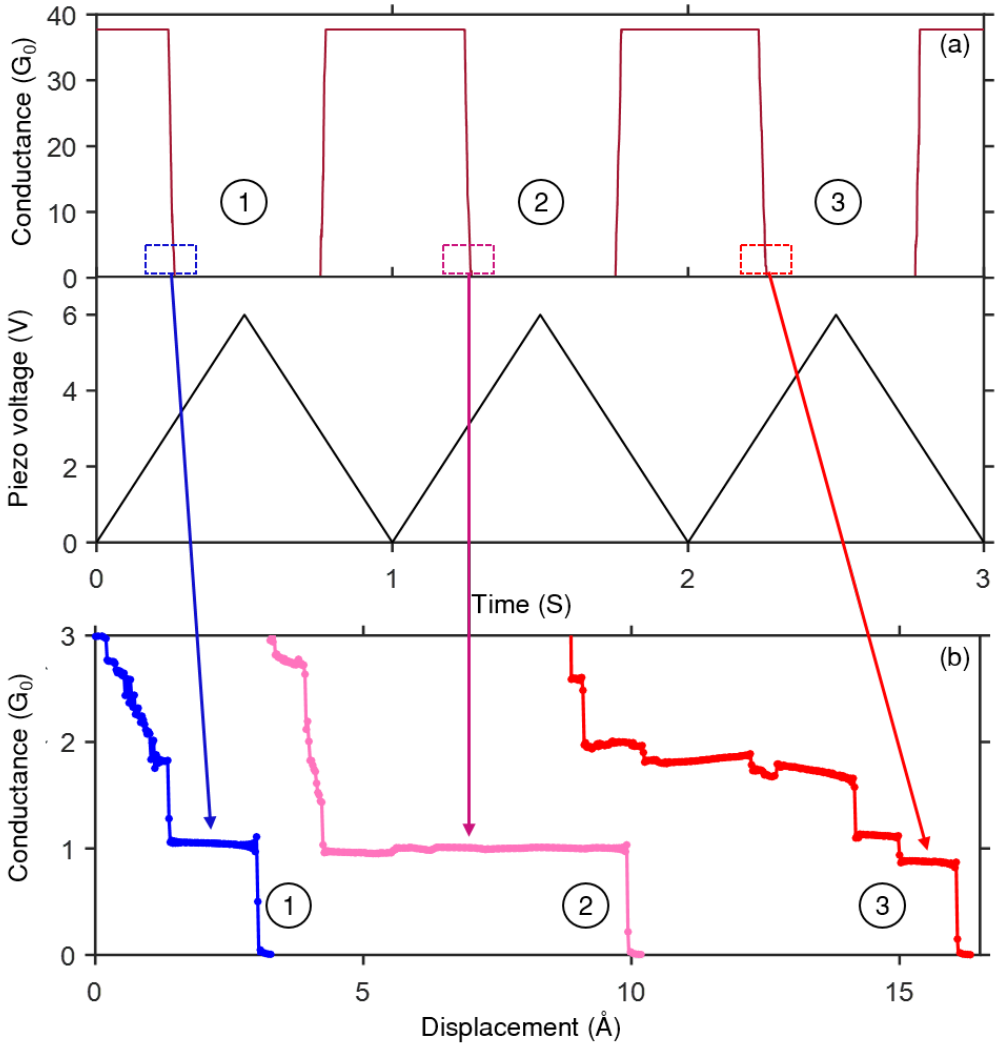

Figure S1: (a) Acquisition of conductance traces for a gold (Au) atomic contact during consecutive breaking (pull) and making (push) in the break junction configuration at 77K. Bottom panel: Triangular waveform of the piezo voltage and Top panel: corresponding conductance of the junction. (b) Zoomed view of the conductance displacement breaking traces of gold atomic junction, shown in (a) where traces are shifted horizontally for better distinction.

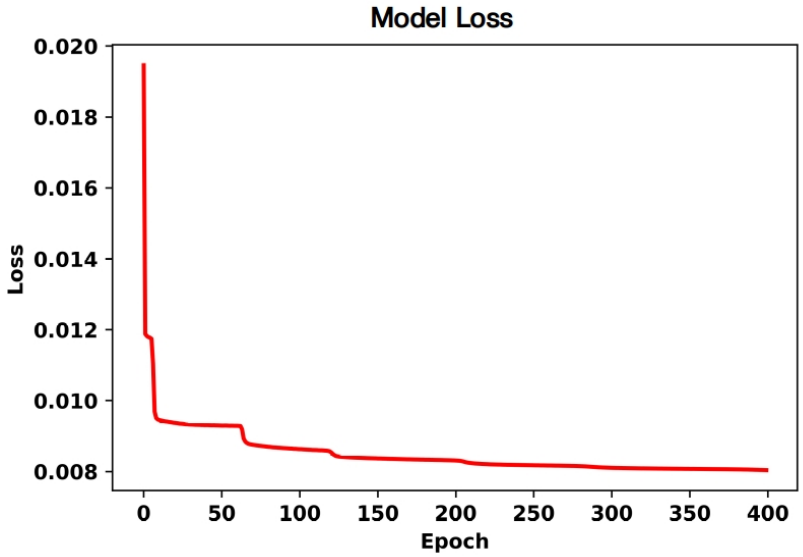

Figure S2: The Loss function during training of the model vs Epoch

S2 Unsupervised Machine Learning Methods

S2.1 Feature Extraction using Encoder-Decoder Model

The initial point of this analysis is the dataset of individual breaking traces obtained from experiment. As a part of data preprocessing, each trace in the dataset are made to have the same dimension. The same number of equispaced datapoints M are picked from each trace between conductance values of $0.7G_0$ and $1.4\ G_0$, the region which encapsulates information of atomic chain formation before breaking. A value of M=100 is seen to provide enough accuracy without computational overhead. Consequently, traces with data-points less than M in the specified range are discarded. This reduces the original dataset to $\sim$ 22000 from 90000. Thus we end up with a modified dataset of the form $\{G_n(m);\ n \in 1,2,..N;\ m \in 1,2,..M\}$ where n is the trace index and m is the points in each trace.

The unsupervised learning workflow adopted here for extracting the essential features from the traces is the Stacked auto-encoder model, shown schematically in figure 1(C) in the main text. This neural network based deep learning framework essentially consists of two parts, an encoding part and a decoding part. The encoding part produces a dimension-reduced feature representation of the input space into a code layer, while the decoding

part reconstructs the initial input using the code as input by minimizing a loss function. The dimension reduction(reconstruction) in the encoder(decoder) part is done through hierarchically stacked hidden layers of neurons where the output of one layer is fed to the input of the next layer. Let L_I and L_F to be the number of neurons in the input data and feature space, respectively. For a set of N training samples of the data space $X = \{x_i \in R^{L_I}\}$ of the data space, the encoding network with parameters P_e maps x_i to $\{y_i \in R^{L_F}\}$ in the feature space(Y) in a non-linear fashion. Mathematically this can be represented as $Y = f^{enc}{}_{P_e}(X)$. The decoding network with parameters P_d is used to reconstruct the data from feature space to the original data space by another set of non-linear mapping. This equates to the formulation $X = f^{dec}{}_{P_d}(Y)$. If the data space has dimension much larger than that of the feature space,i.e, $L_I \gg L_F$, then the outputs of the encoder network represent the latent key features of the input trace data. The encoding-decoding network are optimized with respect to their parameters P_e and P_d by training them on the pre-processed experimental trace data $\{G_n; n=1,2,..N\}$. The training aims at minimizing the loss function(Fig S2) :

$$Loss(P_e, P_d) = \sum_{n=1}^{N} \parallel f_{P_d}^{dec}(f_{P_e}^{enc}(G_n)) - G_n \parallel^2$$

There are certain algorithm-specific hyperparameters that are connected to the training of the stacked autoencoder model. This include the size of the code layer or the feature space, number of layers in the encoder and decoder and the number of nodes in each layer. The detailed network architecture is given in Table S1. Adam optimizer has been used in the model. Additionally, two other parameters required in neural network training, the learning rate and the batch size, are set to 1e-5 and 100 respectively. The model is trained for 400 epochs beyond which the loss function saturates.

S2.2 Feature Clustering and Representation with K-means and PCA

A clustering algorithm is used to distinguish among different kinds of traces based on the extracted features in the previous step. This aids in grouping breaking traces generated through different atomic rearrangements. A K-Means clustering algorithm is applied for our purpose. The choice is made owing to its efficiency and effectiveness in handling higher dimensional data. This is an iterative process to conglomerate data into K different clusters such that within each cluster the sum-squared-distances are minimized from their respective centroids. The process begins by placing the centroids randomly in

	Layer Name	No of Nodes	Activation function
	Input Layer	100	None
Encoder	Hidden Layer 1	75	Sigmoid
	Hidden Layer 2	50	Sigmoid
Feature	Feature Layer	25	Sigmoid
	Hidden Layer 1	50	Sigmoid
Decoder	Hidden Layer 2	75	Sigmoid
	Output Layer	100	None

Table S1: The network architecture of the stacked auto-encoder model

the feature space. Next, distance between each trace and all the centroids are calculated and is subsequently assigned to the cluster of the nearest centroid. The centroids are then recalculated. This cycle is iterated until maximum displacement of the centroids(O_j, where j are centroid indices) between two consecutive iterations converge to a small value(ϵ), i.e, max$\| O_j^{s+1} - O_j^s \| < \epsilon$, where s+1 and s indicate two consecutive iterations.

The choice of the number of clusters(K) is undoubtedly crucial to get clustering that yields coherent physical interpretation. To determine the optimal value of clusters K_{true}, we have implemented the *Elbow Method algorithm*. The idea is to calculate the Sum-of-squared-errors(SSE) for several guess K values. SSE declines rapidly when the number of clusters K ($K<K_{true}$) approach the true number of clusters(K_{true}) in the data, and quickly slows down once it is exceeded. The optimal K_{true} value can then be determined by plotting K-SSE curve and locating the point of inflection[1]. Rather than locating this point visually, kneedle[2] algorithm is implemented via KneeLocator[3] that returns the point of maximum curvature(knee) for a set of (x,y). The optimal number of clusters is found to be three(Fig S3).

For better comprehension of the clustering exercise, the trace data from the multidimensional feature space are projected to a lower dimensional dataspace without loss of crucial information. This is achieved by employing the Principal Component Analysis (PCA)[4] algorithm. In this technique, every vector of the original feature space is projected onto the first two eigenvectors of the covariance matrix of the data.

S3 Chain formation analysis

Mechanically controllable break junction (MCBJ) technique what we have

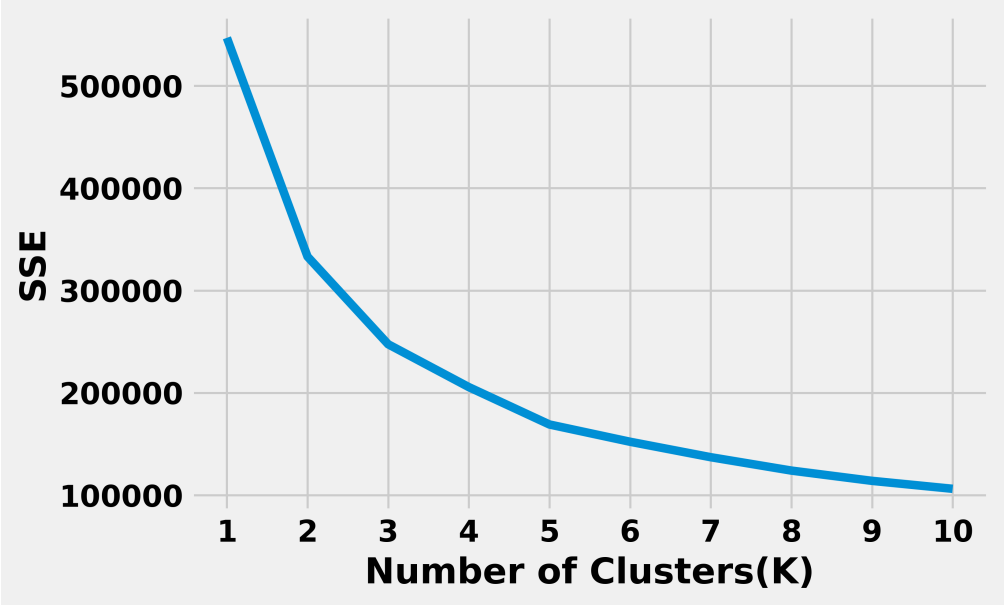

Figure S3: SSE-vs-Number of Clusters(K). The point of inflection lies at K=3

used to characterize the atomic junctions, can provide statistics over thousands of independent data sets, and provides unambiguous evidence of chain formation, as well accepted in the community. In the following, we provide a detailed account of this.

S3.1 Plateau length histogram

Typically, if there is a formation of atomic chain, the plateau length ($1G_0$) will be longer. Technically, it is defined as the difference between two absolute displacement values, corresponding to a conductance G_i to a desired conductance value G_f. Two representative cases are shown having both breaking (blue) and making (red) traces (see Figure S4c-d). Histogram of these length for breaking traces is thus presents the distribution of plateau lengths, obtained from each traces. It has been widely used to probe the atomic chain formation in metals [5, 6, 7, 8]. Figure S4a shows the plateau length histogram of gold atomic junctions where (1), (2) and (3) denotes three consecutive peaks with average peak value 1.67 ± 0.11 Å, 4.78 ± 0.17 Å and 6.95 ± 0.32 Å. Conductance segments used to prepare the histogram is 0.8 G_0 to 1.2 G_0. It is evident that plateau lengths can extend up to 10 Åas well as histogram exhibits equidistant peaks of average separation 2.6

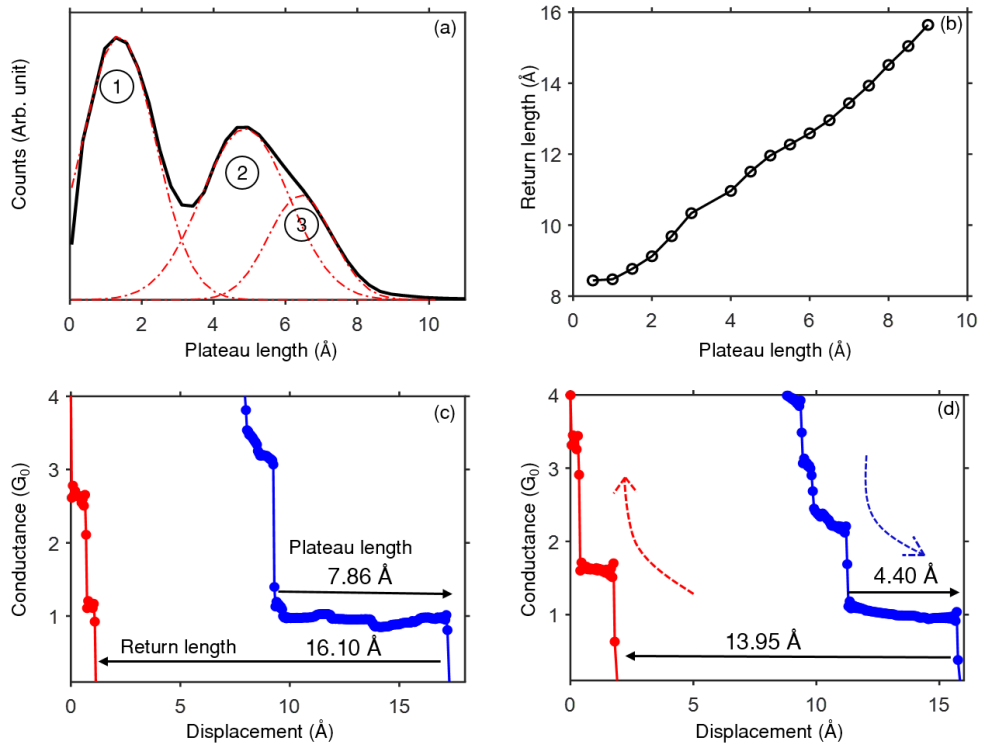

Figure S4: (a) Plateau length histogram of gold atomic junction, measured at 77 K where red dash dotted line denotes the Gaussian fitting of the corresponding peaks: 1, 2 and 3. (b) Average return length vs. plateau length of gold atomic junction, measured at 77 K. (c, d) Two representative conductance displacement breaking (blue) and making (red) traces of gold atomic junction where breaking and making event is denoted by the arrow in d. The black arrows, representing the plateau length and return length with their absolute value, are mentioned in the respective graph.

Å. Length corresponding to each peak structure demonstrates the tendency of the junction to break at that displacement values, in which a sufficient amount of stress is accumulated upon elongation. The relative separation between these peaks is of the order of the interatomic separation of gold, which is an indicative of gold atomic chain formation.

S3.2 Return length histogram

Further confirmation on the formation of atomic chain is established by plotting average return length as a function of plateau length, shown in Figure

S4b. Return length is defined as the length over which two electrodes are needed to retract back to reestablish the junction after breaking. For instance, plateau length and return length of two representative traces are shown in Figure S4c-d. Figure S4b shows the average return length calculated from large number of traces as a function of the corresponding plateau length and we observed that average return length increases almost linearly with the plateau length which gives a strong indication regarding the formation of Au atomic chain. An offset of around 8 Åis observed to be present due to the elastic response of the atomic banks[5, 6, 9, 10].

S4 Supervised Machine Learning Method

Random Forest algorithm has been employed in our work. The number of trees in the forest has been set to 500 which corresponds to the optimal model. No limitations has been imposed on the depth of a tree. Based on 10-fold cross-validation, the optimal model has correlation coefficient of 0.94 and mean-absolute error of ∼3%.

S5 Comparison of 1D histograms

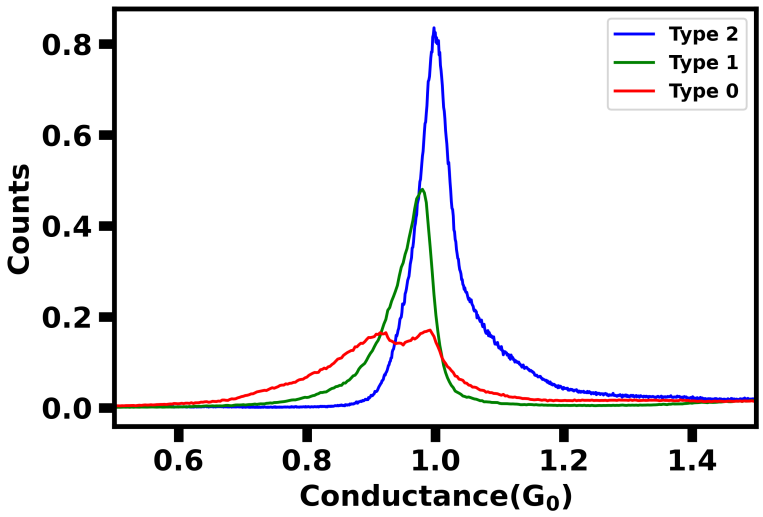

Figure S5: A comparison of the 1D histograms of three types demonstrating the characteristics difference among them.

References

[1] Dhendra Marutho, Sunarna Hendra Handaka, Ekaprana Wijaya, and Muljono. The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In *2018 International Seminar on Application for Technology of Information and Communication*, pages 533–538, 2018.

[2] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, pages 166–171, 2011.

[3] Kevin Arvai. kneed, August 2020.

[4] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52, 1987. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[5] AI Yanson, G Rubio Bollinger, HE Van den Brom, N Agrait, and JM Van Ruitenbeek. Formation and manipulation of a metallic wire of single gold atoms. *Nature*, 395(6704):783–785, 1998.

[6] WHA Thijssen, D Marjenburgh, RH Bremmer, and JM Van Ruitenbeek. Oxygen-enhanced atomic chain formation. *Physical review letters*, 96(2):026806, 2006.

[7] C. Untiedt, A. I. Yanson, R. Grande, G. Rubio-Bollinger, N. Agraït, S. Vieira, and J.M. van Ruitenbeek. Calibration of the length of a chain of single gold atoms. *Phys. Rev. B*, 66:085418, Aug 2002.

[8] RHM Smit, Carlos Untiedt, AI Yanson, and JM Van Ruitenbeek. Common origin for surface reconstruction and the formation of chains of metal atoms. *Physical Review Letters*, 87(26):266102, 2001.

[9] Tomoka Nakazumi, Satoshi Kaneko, and Manabu Kiguchi. Electron transport properties of au, ag, and cu atomic contacts in a hydrogen environment. *The Journal of Physical Chemistry C*, 118(14):7489–7493, 2014.

[10] Sune R Bahn and Karsten W Jacobsen. Chain formation of metal atoms. *Physical review letters*, 87(26):266101, 2001.

[11] Biswajit Pabi, Debayan Mondal, Priya Mahadevan, and Atindra Nath Pal. Probing metal-molecule contact at the atomic scale via conductance jumps. *Phys. Rev. B*, 104:L121407, Sep 2021.

[12] Biswajit Pabi and Atindra Nath Pal. An experimental set-up to probe the quantum transport through a single atomic/molecular junction at room temperature. *Pramana*, 97, Dec 2022.

[13] Rick Bitter, Taqi Mohiuddin, and Matt Nawrocki. *LabVIEW: Advanced programming techniques*. Crc Press, 2006.

[14] Johannes Fürnkranz. *Decision Tree*, pages 263–267. Springer US, Boston, MA, 2010.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[16] David Freedman, Robert Pisani, and Roger Purves. Statistics (international student edition). *Pisani, R. Purves, 4th edn. WW Norton & Company, New York*, 2007.

[17] Payam Refaeilzadeh, Lei Tang, and Huan Liu. *Cross-Validation*, pages 532–538. Springer US, Boston, MA, 2009.

[18] Anita Halder, Aishwaryo Ghosh, and Tanusri Saha Dasgupta. Machine-learning-assisted prediction of magnetic double perovskites. *Phys. Rev. Mater.*, 3:084418, Aug 2019.

[19] Anita Halder, Samir Rom, Aishwaryo Ghosh, and Tanusri Saha-Dasgupta. Prediction of the properties of the rare-earth magnets $ce_2fe_{17-x}co_xCN$: A combined machine-learning and ab initio study. *Phys. Rev. Appl.*, 14:034024, Sep 2020.