Transformer-Based Deep Learning Structure-Conductance Relationships in Gold and Silver Nanowires

Dongying Lin,[1] Jijie Zou,[1,2] Yangyu Dong,[1,2] Yudi Wang,[1*] Yongfeng Wang,[1] Stefano Sanvito,[3] Shimin Hou[1,2*]

1 Key Laboratory for the Physics and Chemistry of Nanodevices, School of Electronics, Peking University, Beijing 100871, China

2 Centre for Nanoscale Science and Technology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

3 School of Physics, AMBER and CRANN Institute, Trinity College, Dublin 2, Ireland

Contents:

1. **The performance of the neural-network potentials on the Au compact nanowire**

We further test the performance of the neural-network potentials (NNPs) developed by Andolina and coworkers on the Au compact nanowire.[1] Taking the starting configuration, the relaxed structure and the one forming a monatomic chain as typical examples, we compute the total energy and the atomic forces employing both the NNP and DFT calculations implemented in VASP. The input parameters for DFT calculations are the same as those used for generating the NNP training database.[1] Considering that Andolina and coworkers have demonstrated the excellent performance of the developed NNP on bulk gold, we remove most of the gold atoms in the left and right electrodes and only keep two monatomic layers sandwiching the central nanowire, as shown in Figure S1. The mean absolute errors on the atomic forces computed using the NNP and DFT calculations are on the order of 10^{-2} eV/Å for these three typical configurations (see Table S1), confirming its applicability to modeling the structural evolution of gold nanowires during the stretching process.

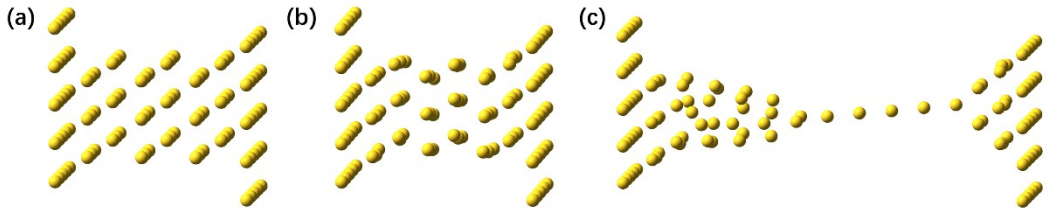

Figure S1 Three typical configurations of the Au compact nanowire used for testing the NNP performance: (a) the starting configuration, (b) the relaxed structure, and (c) the one forming a monatomic chain.

Table S1 The mean absolute of errors (MAE) on the atomic forces computed using the NNP and DFT calculations

	MAE (unit: eV/Å)
The staring configuration	3.58×10^{-2}
The relaxed structure	5.16×10^{-2}
The monatomic-chain configuration	4.18×10^{-2}

2. K-points and basis sets used in the NEGF+DFT calculations

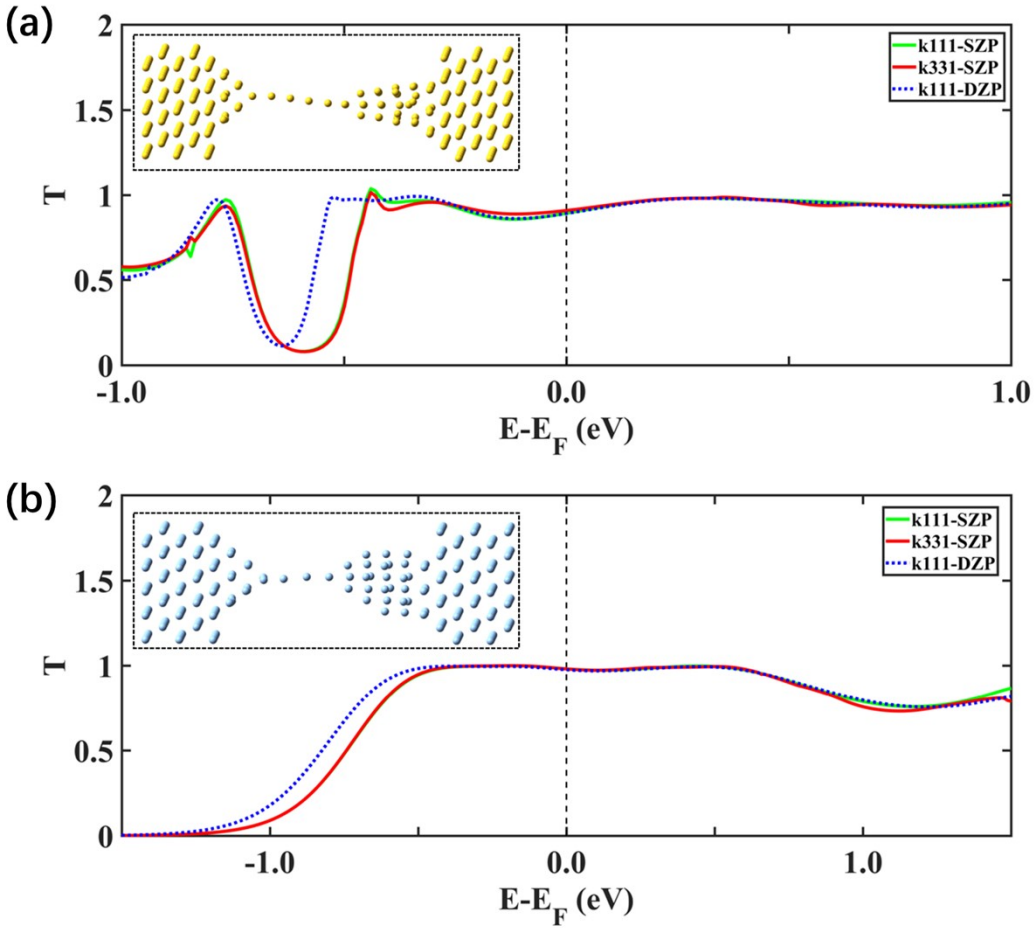

Figure S2 Comparison of the equilibrium transmission spectra of typical snapshots of Au (a) and Ag (b) nanowires calculated with different k-points and basis sets. Inset: atomic structures used to calculate the transmission spectra.

We test the influences of k-points and basis sets on the calculated transmission spectra of gold and silver nanowires. As shown in Figure S2, the transmission around the Fermi level is essentially identical for different k-points and basis sets. Because we are mainly interested in the zero-bias conductance of gold and silver nanowires, which is equal to the product of the conductance quantum $G_0=2e^2/h$ and the transmission coefficient at the Fermi level. Therefore, we choose to use the 1×1×1 k-points and the SZP basis set for our NEGF+DFT calculations considering the computational efficiency.

3. The relation between the zero-bias junction conductance and the transmission coefficient at the Fermi level

As we know, the current-voltage (*I-V*) characteristics of a junction can be obtained from the Landauer formula:[2]

$$I = \frac{2e}{h}\int_{-\infty}^{+\infty} T(E)[f_0(E-\mu_1) - f_0(E-\mu_2)]dE \quad (1)$$

where e is the elementary charge and h is Planck's constant, T(E) is the transmission function of the junction. $f_0(E) = \frac{1}{1+e^{\frac{E}{K_B T}}}$ is the Fermi-Dirac distribution function in which K_B is the Boltzmann constant and T is the temperature. The local quasi-Fermi levels μ_1 and μ_2 of the left and right electrodes are respectively defined as $\mu_1 = E_F + \frac{eV}{2}$ and $\mu_2 = E_F - \frac{eV}{2}$, where E_F is the Fermi level of the junction at equilibrium and V is the applied bias. When a small bias voltage V is applied, the resulting current through the junction can be written to first order as

$$I = \frac{2e^2}{h}V\int_{-\infty}^{+\infty} T(E)[-\frac{\partial f_0(E)}{\partial E}]_{E=E_F}dE \quad (2)$$

so that the conductance of the junction is given by

$$G = \frac{2e^2}{h}\int_{-\infty}^{+\infty} T(E)[-\frac{\partial f_0(E)}{\partial E}]_{E=E_F}dE \quad (3)$$

In the limit of zero temperature, the conductance is reduced to $G=G_0T(E_F)$ where $G_0 = \frac{2e^2}{h}$ is the quantum unit of conductance.

4. Simulated conductance profile of the silver compact nanowire in one single MD trajectory

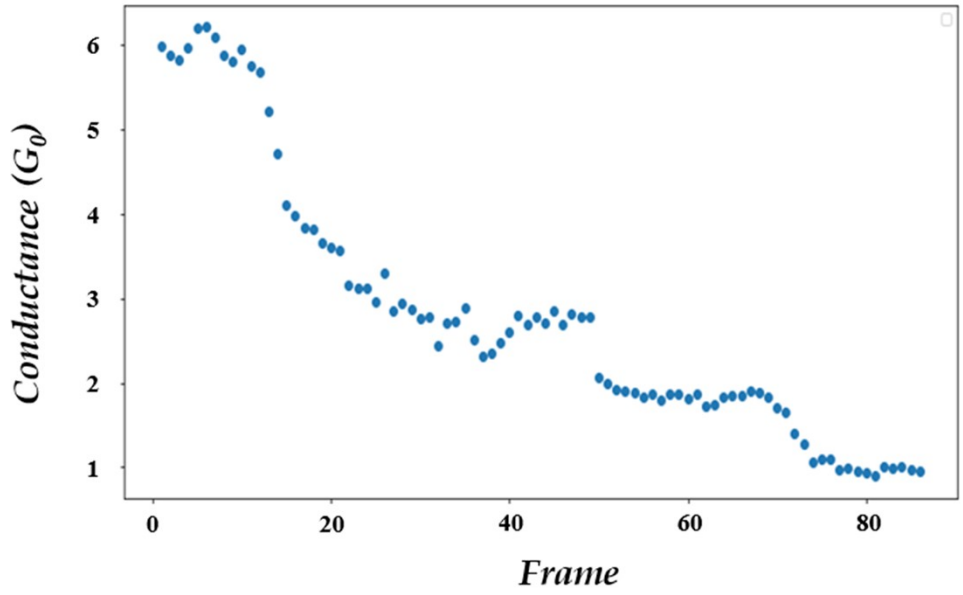

Figure S3 Typical conductance profile during the pulling of the silver compact nanowire.

5. The performance of the fully connected neural network on the prediction of the conductance of Au full nanowires

We have also tested the performance of the fully connected (FC) neural network on the prediction of the conductance of Au full nanowires, and the model architecture of the FC neural network is the same as that employed by Bürkle and coworkers for the prediction of the conductance of potassium nanowires.[3] In contrast to the transformer-based neural network, the performance of the FC neural network strongly depend on the setting of hyperparameters. When the learning rate is greater than 3×10^{-4}, the FC neural network cannot converge well. The learning curve with the learning rate of 5×10^{-4} is shown in Figure S4(a), from which we can see the loss function fluctuates heavily.

We have selected many intermediate models at different training steps and tested their performance on the prediction of the conductance of Au full nanowires. Simulated conductance histograms delivered by three of them with relatively good performance are shown in Figures S4b-S4d, in which the conductance values of MD snapshots of one widened and three elongated gold full nanowires (28,000 structures for each starting structure) are predicted using these three models. Obviously, the consistency among these simulated conductance histograms is worse than that predicted with the transformer-based neural network (see Figures 4c and 4d). Even for the best intermediate FC model obtained at the training step of 2500, the MAE values on the test set of the Au compact wire and a small dataset of the gold full wire with an increased length of 2.29 nm are respectively increased to be 0.09 G_0 and 0.15 G_0, both larger than those (0.06 G_0 and 0.09 G_0) given by the transformer-based neural network.

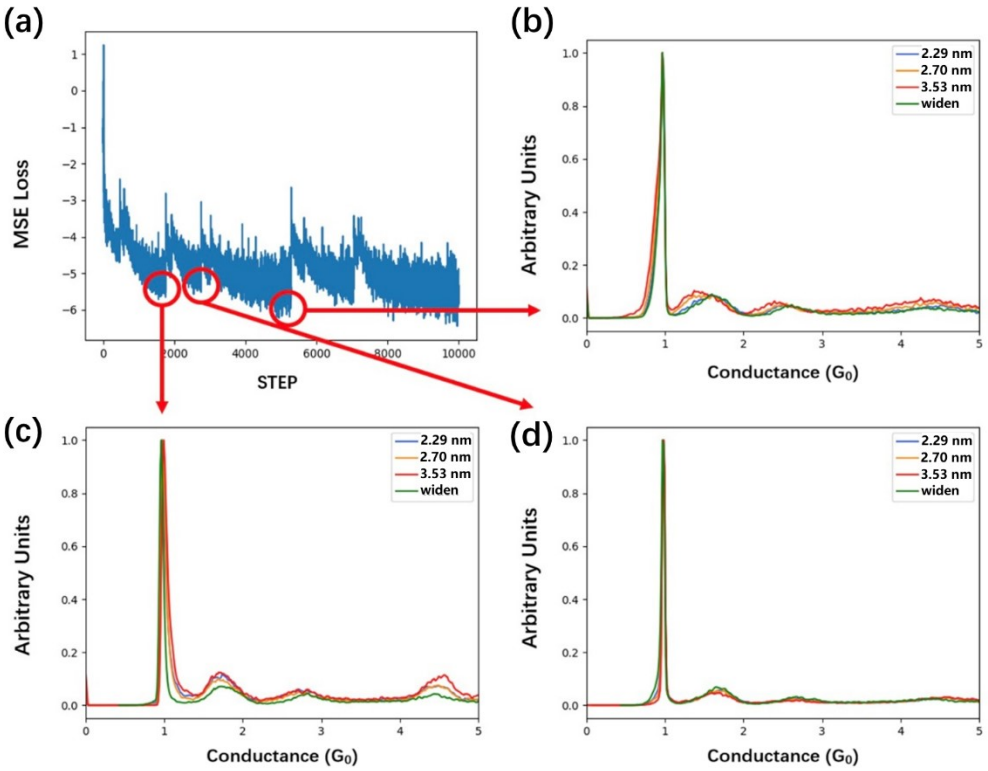

Figure S4 (a) Learning curve of the training loss of the FC model with a learning rate of 5×10^{-4}, the vertical scale is logarithmic. Simulated conductance histograms constructed using the conductance values predicted for three longer and one wider Au full nanowires using the intermediate FC models at the steps of 5000 (b), 1800 (c) and 2500 (d). The lengths of these three elongated Au full wires

are 2.29 nm (blue line), 2.70 nm (yellow line) and 3.53 nm (red line), respectively. The widened Au full nanowire consists of 6 monatomic layers with a 4×4 in-plane supercell (green line).

Then we decrease the learning rate to 1×10^{-4}, and the learning curve and the simulated conductance histograms are shown in Figure S5. As we can see, the model can converge well; unfortunately, the prediction results become much worse. It seems that the FC model is caught in a local optimality. When the learning rate is further decreased to 1×10^{-5}, the model performance is even worse.

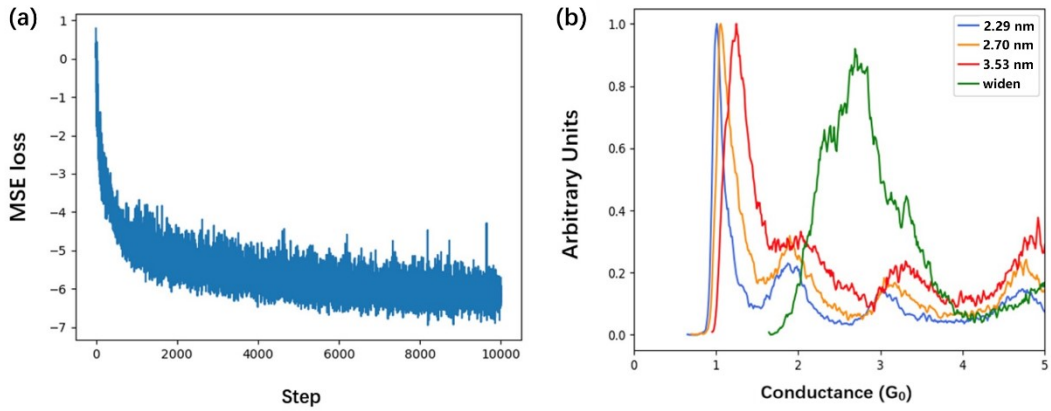

Figure S5 (a) Learning curve of the FC model trained with a learning rate of 1×10^{-4}, the vertical scale is logarithmic. (b) Simulated conductance histograms constructed with conductance values of four Au full nanowires predicted using the FC model trained with a learning rate of 1×10^{-4}.

In contrast, the transformer-based neural network is very stable. As shown in Figure S6, the model converges very well at the learning rate of 1×10^{-5} and the simulated conductance histograms for all the Au full nanowires are also consistent (see Figures 4(c) and 4(d)), demonstrating the robustness of the transformer-based neural network.

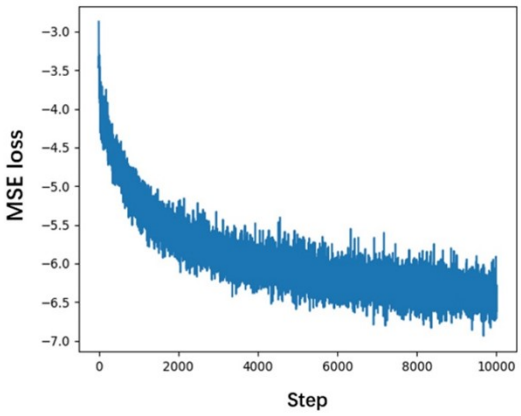

Figure S6 Learning curve of the training loss of the transformer-based neural network with a learning rate of 1×10^{-5}, the vertical scale is logarithmic.

Reference

1. Andolina, C.M.; Bon, M.; Passerone, D.; Saidi, W.A. Robust, Multi-Length-Scale, Machine Learning Potential for Ag–Au Bimetallic Alloys from Clusters to Bulk Materials, *J. Phys. Chem. C*, **2021**, 125, 17438-17447

2. Datta, S. Quantum Transport: Atom to Transistor, Cambridge University Press, 2005.

3. Bürkle, M.; Perera, U.; Gimbert, F.; Nakmura, H.; Kawata, M.; Asai, Y. Deep-Learning Approach to First-Principles Transport Simulations, *Phys. Rev. Lett.* **2021**, 126, 177701