Nuclear Engineering and Technology 57 (2025) 103271

Contents lists available at ScienceDirect

Nuclear Engineering and Technology

journal homepage: www.elsevier.com/locate/net

Original Article

Machine learning study of universal electronic stopping cross-sections of ions in matter

Fan Cheng [a,b], Xun Liu [b,c], Qirong Zheng [a,b], Chuanguo Zhang [a], Bo Da [c,**], Yonggang Li [a,b,*]

[a] Key Laboratory of Materials Physics, Institute of Solid State Physics, HFIPS, Chinese Academy of Sciences, Hefei, 230031, People's Republic of China
[b] University of Science and Technology of China, Hefei, 230026, People's Republic of China
[c] Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Namiki 1-1, Tsukuba, Ibaraki, 305-0044, Japan

ARTICLE INFO

Keywords:
Electronic stopping cross-section
Machine learning
Least absolute shrinkage and selection operator (LASSO)

ABSTRACT

Accurate electronic stopping cross-section (ESCS) database of ions in matter is crucial for precise simulation of radiation damage. Based on the experimental-cleaned database of SRIM, binary theory and unitary convolution approximation as well as the descriptor pool extracted from these models, we developed a universal machine learning ESCS database using the least absolute shrinkage and selection operator (LASSO) algorithm. This method allows for predictions for ion-target combinations with atomic numbers from 1 to 92, within the energy range from 1 keV/u to 1 GeV/u, addressing the limitations of machine learning on training dataset. The database exhibits remarkable accuracy in predicting ESCS and ion depth distribution/range, along with robust reciprocity performance. Key descriptors are also determined, which closely mimic the Lindhard-Scharff-Schiott and Bohr-Bethe-Bloch formulations, achieved through precise adjustments of the exponent of individual elements. The proposed universal ESCS database surpasses the accuracy of existing databases, supporting related applications across a wide range of energies and systems.

1. Introduction

Research on radiation effects in materials is imperative for developing structural components in nuclear fusion reactors, semiconductor devices in aerospace, ion beam-based cancer therapy and so on [1–3]. It is thus important to understand the fundamental effects of energetic particles like ions in matter, especially for the energy deposition of dissipative processes under irradiation. The energy deposition has a direct effect on the primary radiation damage when ions pass through a material. Typically, the energy loss (dE) per unit length (dx) is defined as the stopping power, $-dE/dx$ or $N\sigma(E)$, where N is the atom number density and σ is the stopping cross-section. When ion speeds are far below the velocity of light, the total stopping power can be approximated as the overall contributions of two main independent collision events that, $(-dE/dx)_{\text{total}} = (-dE/dx)_{\text{e}} + (-dE/dx)_{\text{n}}$ [4]. Here $(-dE/dx)_{\text{e}}$ and $(-dE/dx)_{\text{n}}$ represent the electronic and nuclear stopping powers related to the inelastic and elastic interactions of a projectile and the orbital electrons and nuclei of a target atom, respectively. Typically, $(-dE/dx)_{\text{n}}$ is orders of magnitude lower than $(-dE/dx)_{\text{e}}$, becoming noticeable only for large ion masses and/or at very low energies [5]. During ion implantation, ions will lose energy through colliding with the nucleus and electrons of target atoms and finally stop in matter, influencing the electrical and mechanical properties of materials. Inaccuracy in calculating electronic stopping power can cause errors up to 30 % in ion range calculations [6].

Over the past century, many experimental and theoretical electronic stopping powers have been obtained with the rapid development of atomic physics. It is very hard for experimentalists to measure electronic stopping powers over a wide energy range from eV to GeV for all ion-target combinations. Most data are given for light ions like H and He, while scarce data for heavy ions are mostly concentrated near the Bragg peak [7]. Theoretical and empirical models of electronic stopping power are mainly developed by describing the electronic stopping cross-section (ESCS) with five basic descriptors of the atomic number (Z_1, Z_2) and

* Corresponding author. Key Laboratory of Materials Physics, Institute of Solid State Physics, *HFIPS*, Chinese Academy of Sciences, Hefei, 230031, People's Republic of China.
** Corresponding author.
E-mail addresses: DA.Bo@nims.go.jp (B. Da), ygli@theory.issp.ac.cn (Y. Li).

https://doi.org/10.1016/j.net.2024.10.033
Received 29 July 2024; Received in revised form 7 October 2024; Accepted 15 October 2024
Available online 22 October 2024
1738-5733/© 2025 Korean Nuclear Society. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

mass number (M_1, M_2) of projectile ions and targets, and ion incident energy (E_i) [8]. In 1913, Bohr developed the first theoretical model of ESCS based on the classical atomic model [9]. Then Bethe developed a quantum perturbation theory that is completely different from the Bohr theory [10]. Bloch connected Bohr and Bethe's results with the semi-classical theory for the free-electron gas [11]. These models (B-B-B) provide the general framework to almost all contributions on the energy loss of swift particles in matter [12]. For the low energy range ($v < v_0$, where v_0 is the Bohr velocity), the Lindhard-Scharff-Schiott (LSS) formula provides an approximation solution for the calculation of ESCS by approximately considering the interaction of ion and electron as a charged particle with the free-electron gas [13]. Based on the B-B-B theories and LSS-based formulations, many theoretical and semi-empirical models are proposed. The most commonly used models include Stopping and Range of Ions in Matter (SRIM) [14], Binary Theory (BT) [15], Perturbation/Unitary Convolution Approximation (PCA/UCA) [16], Shell wise Local Plasma Approximation (SLPA) [17] and MSTAR [18]. Among them, SRIM, BT and UCA are typical theoretical ESCS with the most comprehensive database. SRIM collects the limited available experimental results, with assuming that the Bethe-Bloch theory served as a guide at the whole energy range [14,19]. BT is an extension of the Bohr's classical theory that considers the interaction of a swift point charge with independent electrons bound harmonically to target nucleus [15]. UCA is based on the Bloch's non-perturbation theory, and calculates the impact-parameter dependence electronic energy loss from the convolution of the target electronic density integrated along the ion path [16]. Whilst present understanding of particle energy loss in matter is vast, it is hard to tell which model is the best option for specific projectile ion and specific energy. An accurate ESCS database is still required that combines the advantages of these models for any materials and projectile ions, and across a complete energy range.

The machine learning method can be trained to identify patterns in example data and accurately generate predictions for unseen data [20] and has been widely used in the prediction of ion/electron electronic stopping [5,8,21–23] and inelastic mean free path [24]. This enables the possibility of learning from the existing dataset to directly predict precise ESCS. A comprehensive and accurate ESCS database should meet three criteria: First, the data cover a wide energy range from 1 keV/u to 1 GeV/u for ion-target combinations of atomic numbers from 1 to 92, with good versatility. Secondly, the predicted data should be consistent with the existing experimental ones. Finally, the database can be used to accurately simulate primary radiation damage comparable to experiments or atomistic models like MD. Many attempts have been made to train with machine learning in recent years. The machine learning algorithms including the random forest, deep learning and stacking methods are used by training the experimental data collected by the International Atomic Energy Agency (IAEA). The predicted ESCS can agree with experiments in most cases [5,8,21,23]. However, there are three limits in these researches: First, the predicted ESCS strongly depends on experimental data, which show discontinuities in energy ranges with scarce experimental data especially for heavy ions. Secondly, the results predicted are discrete in low energies, deviating from the linear relationship of the LSS formula. Finally, the ESCS are trained using only 5 basic descriptors, meaning that only the total ESCS can be predicted without detailed key descriptors. Therefore, a proper advanced machine learning-based model surpassing these classical methods is expected to extract the advantages of existing ESCS data and theoretical models.

In this work, theoretical ESCS data is used as the training dataset instead of experimental data, overcoming the limitations of current machine learning applications in ESCS training. A universal ESCS database that meets the above three criteria is obtained by training the experimental-cleaned theoretical data. An extensive array of features derived from established ESCS theories is incorporated in the training, with key features identified using the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm [24]. The validation using experimental data [25,26] in-depth analysis of primary radiation damage results simulated by IM3D [27], and a thorough examination based on the reciprocity theory [28] to ensure that the universal ESCS database serves as a robust resource for the simulation of primary damage processes.

2. Theoretical model

2.1. Machine learning scheme with LASSO

In machine learning, the linear regression method has the advantage of good interpretability. As a kind of linear regression, LASSO is an ideal method to train the high-dimensional sample dataset and has the superiority of extracting the most critical descriptors included in various theoretical models as shown in Fig. 1. It provides low-dimensional solutions by recasting a problem into a convex minimization problem. A sharp reduction in the number of descriptors is mathematically achieved by solving a minimization function,

$$J_{\min} = \min_{\beta} \frac{1}{2n_{\text{samples}}} \|X\beta - y\|_2^2 + \alpha \|\beta\|_1, \tag{4}$$

where X and y are the target and the predicted values, respectively, n_{samples} is the number of samples. The penalty function contains a constant α (0.0001 for both low and high energies here) and the coefficient vector $||\beta||_1$. The use of $||\beta||_1$ is key for reducing descriptors and avoids the model being overfitted. LASSO used here is driven by the Scikit-learn library [29]. The details of LASSO can be found elsewhere [30].

The ESCS training dataset contains the data of ion-target combinations with atomic numbers from 1 to 92 and in the energy range of 1 keV/u - 1 GeV/u. The target values of ESCS span 8 orders of magnitudes from 10^{-3} to 10^{5} (in the unit of 10^{-15} eV cm^2/atom), resulting to a big bias in the training. As given in Eq. (4), it is easier to minimize the loss function by prioritizing data with larger values. Therefore, the smaller values of ESCS would be ignored, leading to inaccurate predicted ESCS of light ions or light targets in the low energy range. Hence, the logarithm of the target values is used here to make sure that the magnitude of ESCS is within an order.

2.2. Datasets

The ESCS is trained in the energy range of 1 keV/u - 2 MeV/u and 1 MeV/u - 1 GeV/u, by using the LSS and modified formula, and the B-B-B theory, respectively. The LSS formula is assumed to be valid up to $v_0 Z_1^{2/3}$. The overlap of the low- and high-energy range (1 MeV/u - 2 MeV/u) facilitates the connection of their predictions, which will be described in detail in section 2.4.

The training dataset is composed of SRIM (σ_e^{SRIM}), BT (σ_e^{BT}) and UCA (σ_e^{UCA}), and cleaned with experimental data [25,26]. Data for SRIM is

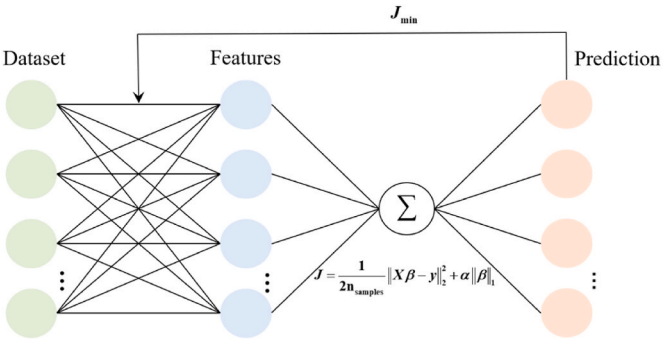

Fig. 1. Framework of the machine learning process of ESCS using the LASSO algorithm.

calculated based on the material physical state at room temperature and BT, UCA both distinguish the target physical state in the model. UCA is calculated with mean charge state while BT with equilibrium charge. Thus the material physical state and ion charge state effects are both included in the training dataset. The three theoretical databases contain the ESCS of ion-target combinations with atomic numbers from 1 to 92 and in the energy range of 1 keV/u-1 GeV/u, except for SRIM in a wider energy range of 1 eV/u-2 GeV/u. In the low energy range, the upgraded SRIM database, BT with the shell correction for light ions, and UCA with the Bloch correction for heavy ions are concluded in Ref. [7]. In the high-energy range, the three theoretical models without cleaning are all used as the training dataset. Here we take Kr ions incident into Al target as an example, as shown in Fig. 2. In the low energy range, the three theoretical ESCS and the LSS theory show a linear relationship with the incident energy, and show great difference with each other. SRIM and BT are in better agreement with the experimental values, so they are selected as the training dataset. In the high energy range, the three theoretical ESCS are consistent with the B-B-B theory and the experimental data so all the three theoretical ESCS are selected as the training dataset. For ion-target combinations with atomic numbers from 1 to 92, 561779 and 390484 theoretical data in total are used as the training datasets in the low- and high-energy ranges, respectively.

2.3. Descriptor pool

According to the ESCS theoretical studies over the past century, the descriptor pools are constructed separately for low- ($E < 2$ MeV/u) and high-energy ($E > 1$ MeV/u) ranges. At low energies, the descriptors can be selected by analyzing the LSS formula [13],

$$S_e = \xi 8\pi e^2 a_0 N \frac{Z_1 Z_2}{Z}\left(\frac{v}{v_0}\right), \tag{1}$$

and its correction formulas [31–33]. Here $Z = \left(Z_1^{2/3} + Z_2^{2/3}\right)^{3/2}$ and $\xi \cong Z_1^{1/6}$. e, a_0 and N are the elementary charge, Bohr atomic radius and atomic density, respectively. v is the incident ion velocity.

There are several corrections to the LSS formula. One of corrections is the modification of Z. Sugiyama replaced Z with the screening term Z_s in the Hartree-Fock-Slater screening function that [31],

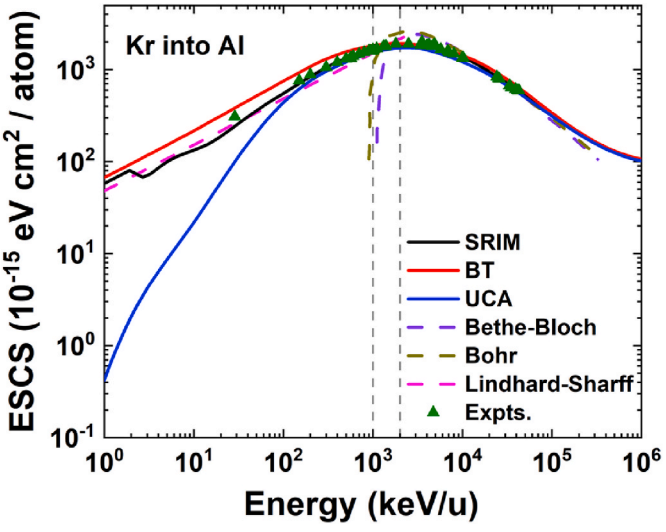

Fig. 2. Comparison of the experimental and theoretical ESCS of Kr ions into Al target.

$$Z_S = \left\{ \begin{array}{l} Z_{1S} + Z_{1S}, Z_1 \approx Z_2 \\ \frac{1}{2}\left(Z_{1S}^{2/3} + Z_{2S}^{2/3}\right)^{3/2}, \text{low } Z_1 \text{ and } Z_2 \\ \frac{1}{2}\left(Z_{1S}^{1/2} + Z_{2S}^{1/2}\right)^{2}, \text{other factors} \end{array} \right\}. \tag{2}$$

The second correction is the fitting of the key parameters in Eq. (1), such as the correction factor ζ and Z_1. This method is only valid for limited systems, such as Mg, Al and Si ions crossing Formvar foil. Two typical modifications of ζ are, $\zeta = a(Z_1)^b\left(\frac{v}{v_0}\right)^c (Z_2)^d$ and $\zeta = a\left(Z_1^b\, Z_2^c\right)$ [32,33], respectively. Here the parameters of a, b, c and d are fitted with the experimental values. The effective charge Z_1^{eff} of moving heavy-charged particles penetrating into matter at the velocity range of $v < v_0 Z_1^{2/3}$ is given as $Z_1^{\text{eff}} = (1 - \exp(-1.25x))Z_1$ instead of Z_1 in Eq. (1), where $x = 100\beta/Z_1^{2/3}$ and $\beta = v/c$ [31]. Other corrections are the shell and Barkas corrections that consider the intrinsic electron motion of each shell and the high-order Barkas effects [34] that, $\Delta L_{\text{shellcor}} \propto Z_2^{4/3} v_0^2/v^2$, and $\Delta L_{\text{Barkas}} \propto Z_1 Z_2 v_0^3/v^3$, respectively. The starting point for building the descriptor pool is the seven basic descriptors of Z_1, Z_2, Z_{1S}, Z_{2S}, M_1, M_2, and E_i, shown in Line A in Table 1. The descriptors in Lines A1, A2, A3 and A4 are the power exponents of the basic descriptors appeared in Eqs. (1) and (2). $\ln Z_2$ and $\ln M_1$ is considered in Line A5. Line B represents the Barkas and shell corrections. Lines C and D lines represent to the main function of the LSS formula and its modification formulas mentioned above. In Line C, the modification of $Z = 1/2\left(Z_{1S}^{1/2} + Z_{2S}^{1/2}\right)^2$ is chosen that is suitable to the more ion-target systems. In total, the descriptor pool for the low energy range contains 895 descriptors.

For high-energy range, the B-B-B model [9–11] is typically used for calculating the ESCS. Based on the assumption that the ion lost all electrons, the ESCS is characterized as the relationship with the ion velocity as,

$$-\frac{dE}{dx} = \frac{4\pi Z_1^2 Z_2 e^4}{mv^2} NL, \tag{3}$$

where

$$L = \left\{ \begin{array}{l} \ln\frac{2mv^2}{\hbar\omega}, \text{Bethe} - \text{Bloch} \\ \ln\frac{Cmv^3}{Z_1 e^2 \omega}, \text{Bohr} \end{array} \right\}. \tag{4}$$

where ω is an effective resonance frequency of the target electrons and $C = 1.1229$ [34]. The basic descriptors of Z_1, Z_2, M_1 and E_i are extracted from Eqs. (3) and (4). The first step is listing the power exponents of Z_1, Z_2, and E_i, as shown in the first three lines of Table 2. Then, following the format of Eq. (3), descriptors containing the logarithmic terms are included, such as $\ln Z_1$, $\ln M_1$ and $\ln E_i$. These

Table 1
Descriptor combination based on five basic descriptors for low-energy range.

ID	Descriptor	Number of descriptors
A	Z_1, Z_2, Z_{1S}, Z_{2S}, M_1, M_2, E_i	7
A1	Z_1^i, i = −1/2, 2/3, 7/6, 2, 5/2, $\ln Z_1$	6
A2	Z_2^i, i = −1, −2, 2/3, 6/5	4
A3	E_i^i, i = 1/2, −3/2, −5/2, ln*E*i	4
A4	Z_{1S}^i, Z_{2S}^i, i = 0.5	2
A5	$\ln Z_2$, $\ln M_1$	2
B	A1·A2·A3	96
C	$\left(Z_1^{0.7} + Z_2^{0.7}\right)^{3/2}$, $\left(Z_{1S}^{0.5} + Z_{2S}^{0.5}\right)^2$	2
D	$\exp(-i \cdot E_i)$, i = 3,4,0, $\exp\left(-125 \cdot E_i^{0.5} \cdot Z_1^{-2/3}\right)$	4
E	B·C·D	768
Total	#	895

Table 2
Descriptor combination based on five basic descriptors for high-energy range.

ID	Descriptor	Number of descriptors
A	Z_1, Z_2, M_1, M_2, E_i	5
A1	Z_1^i, i = 1/5, 1/2, 2/3, 7/6, 3,0, −2, −1	8
A2	Z_2^i, i = 2/3, 1/2, 2, 7/3, lnZ_2	5
A3	E_i^i, i = 3/2,1/2, −1, −2, −5/2, 0	6
A4	lnZ_1, lnM_1, lnE_i, 1	4
A	A1·A2·A3·A4	960
Total	#	988

descriptors are linearly combined to consist of the final 988 descriptors for high-energy range.

2.4. Smooth connection of ESCS between low- and high-energy ranges

The Gaussian amplitude fitting and the spline interpolation are applied to connect the data in the two energy ranges. The incident energy E is operated logarithmically as $\log_{10}E$. The relation of ESCS and $\log_{10}E$ is shown as a single peak function. The Gaussian amplitude fitting has the advantage of fitting the function with single peak. It can fit the function well around the Bragg Peak with an accuracy of higher than 95 %.

The Gaussian amplitude fitting is used to produce a set of fitting data in the whole energy range according to the machine learning data with the fitting score higher than 0.99, as shown in Fig. 3. The Gaussian fitting and machine learning data are divided into three parts of A, B and C. The beginning and ending coordinates of part B are determined by the minimum difference in ESCS between the machine learning and Gaussian fitting data in the ranges of 1 keV/u-1 MeV/u and 1 MeV/u-1 GeV/u, respectively. Parts A and C use the machine learning data and part B uses the Gaussian fitting data to smooth the connection in the coincident energy range. The ESCS data over the whole energy range is formed as shown by the blue dash-dot line of Fig. 3. It can be seen that the connected data is smoother compared to the original data. The Gaussian fitting data is only used to smooth the connected data in energy region near the Bragg Peak with only 2%–3% deviation. When an obvious inflection point appears at the conjunction of A and B or B and C, the spline interpolation is extra used to achieve a further smooth connection.

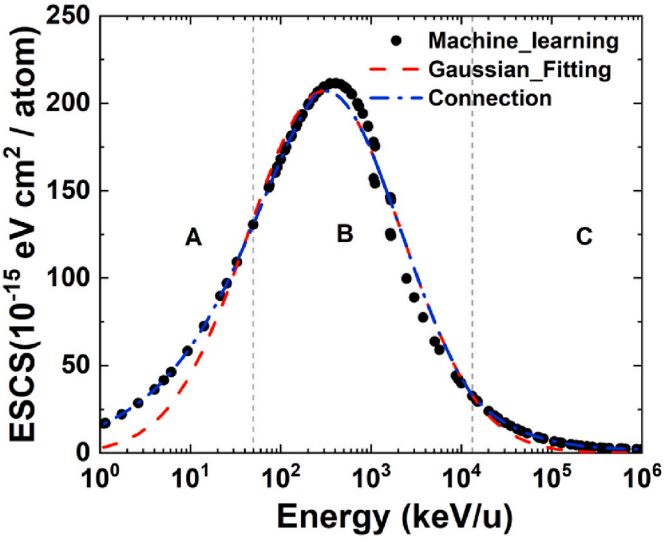

Fig. 3. The smoothed ESCS of Boron ions into Si in different energy ranges with the Gaussian and spline connection.

2.5. Testing and verification

2.5.1. Direct comparation with experimental ESCS

The experimental datasets collected by Paul and IAEA are used as testing dataset, which contains nearly all the published electronic stopping power data of different ions in solids, gases and compounds since the 1990s [25,26]. The testing datasets contain 24784 and 7921 data in the low- and high-energy ranges, respectively. The machine learning predicted ESCS can be directly compared with experiments firstly.

2.5.2. Comparison simulated ion depth distribution with experiments

We introduced the machine learning-based ESCS into our IM3D [27] code to simulate primary radiation damage and compare with experiments. IM3D is a massively parallel, open-source, 3D Monte Carlo code for simulating the transport of ions and the production of defects within materials. We further developed the ESCS module in IM3D for the simulation of primary radiation process with arbitrary ESCS including our machine learning ESCS [35].

The electronic energy loss and nuclear energy loss are commonly treated as two independent processes when predicting ion ranges in the binary-collision approximation (BCA) models. Based on the classical BCA approximation and conserving momentum and energy, the nuclear energy loss can be calculated with the formula of $\Delta E_n = \left(4M_1M_2E\sin^2\theta_{CM}\right)/(M_1+M_2)$, where θ_{CM} is the classical binary atomic scattering angel for elastic collision process and determined by the interaction potential between projectile and target atoms [27].

2.5.3. Reciprocity theory assessment

Sigmund proposed the reciprocity theory by analyzing experimental results [28]. It means the invariance of inelastic excitation in ion-atom collisions against the interchange of projectile and target. Thus, for low velocities slower than Bohr's velocity (about 25 keV/u), it follows *ESCS* (A ion into B target) = *ESCS* (B ion into A target) [28]. Reciprocity is well obeyed with the errors <10 % for most gas, metal and insulator systems. In order to give a quantitative analysis of the reciprocity theory, Guo et al., defined the average relative coherent coefficient (γ_{rec}) based on the reciprocity that [8],

$$\overline{\gamma_{rec}} = \left\langle \frac{|ESCS(E_{in}) - ESCS'(E_{in})|}{0.5|ESCS(E_{in}) + ESCS'(E_{in})|} \right\rangle_{E_{in}}, \quad (5)$$

where $ESCS(E_{in})$ and $ESCS'(E_{in})$ represent the ESCS of A ion incident into B target and the ESCS of B ion into A target, respectively. Both A and B ions have the same kinetic energy E_{in} (in unit of keV/u). γ_{rec} represents the average of the absolute difference between $ESCS(E_{in})$ and $ESCS'(E_{in})$ for E_{in} lower than 25 keV/u.

2.5.4. Error metrics

In assessing the performance of a model, two error metrics are employed to quantify the deviation between the predicted values (y_{pred}) and the experimental reference values (y_{true}). One is the coefficient of determination $R^2 = 1 - \left(\sum\left(y_{true} - y_{pred}\right)^2 / \sum\left(y_{true} - \overline{y_{true}}\right)^2\right)$. Its numerator represents the discrepancies between the model's predictions and the reference values. The denominator represents the discrepancies between the experimental reference value and the mean experimental reference value $\overline{y_{true}}$). The R^2 metric is widely acknowledged to evaluate the fitting of a regression model to the observed data.

3. Results and discussion

3.1. Prediction of ESCS

The theoretical ESCS datasets (SRIM, BT and UCA) of low-energy (1 keV/u - 2 MeV/u) and high-energy (1 MeV/u - 1 GeV/u) ranges are

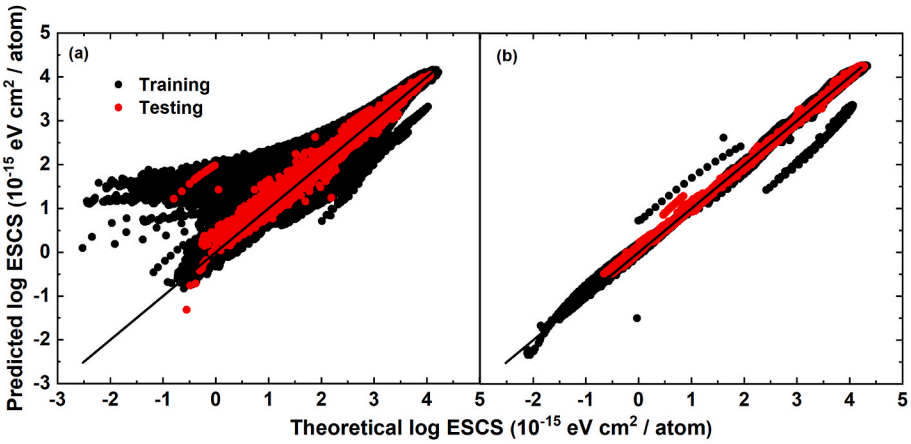

Fig. 4. Predicted ESCS for all samples plotted against the training values in the (a)low- and (b) high-energy ranges, respectively.

trained, respectively. In the low energy range, the training and testing score of R^2 are 0.89 and 0.88, respectively. In fact, the higher deviations or lower training score in the low energy range only introduced 2.50 % of the total training data. In the high energy range, the training and testing score of R^2 both reach up to 0.99, as shown in Fig. 4(b). The machine learning predicted ESCS σ_e^{ML} shows good agreement with the training dataset. Only 0.04 % of the total data show the deviation with the training points.

In the low- and high-energy ranges, the MAPE errors of different systems between the test and σ_e^{ML} are shown in Fig. 5. The MAPE error of all the systems in the low- and high-energy ranges are 12.60 % and 6.20 %, respectively. In the IAEA experimental dataset, the ESCS of H, He, and Li ions account for 55.60 % in the total dataset. Since unpredictable deviations between the data measured by different groups using different instruments in different years, the MAPE errors of light ions is bigger than that of heavy ions in the low energy range.

The comparison of σ_e^{ML} in the full energy range (1 keV/u – 1 GeV/u) with the experimental and theoretical values of different systems is shown in Fig. 6. For each system, σ_e^{ML} can give an accurate prediction with high consistency compared with experimental results. For each system, the theoretical ESCS values and σ_e^{ML} agree well with the experimental ones in the high-energy range, but there are deviations between them in the low-energy range. For low-to-medium energy range, σ_e^{ML} can be consistent with the experimental results and σ_e^{SRIM}. It means that σ_e^{ML} with the accuracy no less than σ_e^{SRIM}. In addition, σ_e^{ML} is better than both

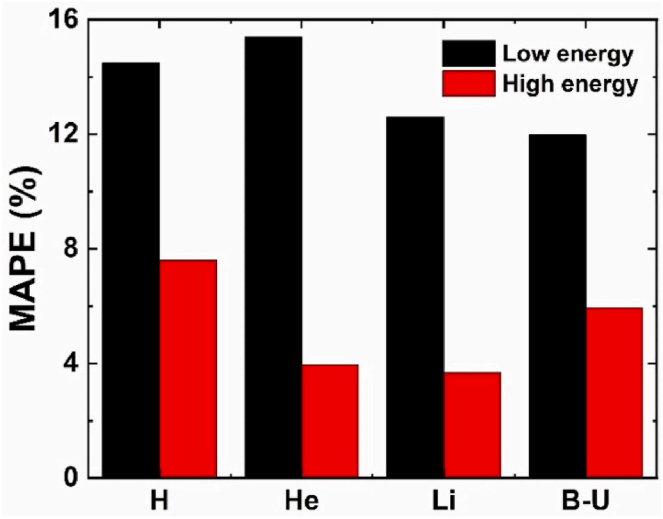

Fig. 5. Alternative MAPE values at low- and high-energy ranges.

σ_e^{BT} and σ_e^{UCA} for H-C, He-W, and B-Au systems. Moreover, in recent years, SRIM has been noted in recent years for its overestimation of heavy ion [6], resulting in an underestimation of ion depth distribution. In Fig. 6(e) of Au ion into C, σ_e^{ML} are lower compared to σ_e^{SRIM}. The machine learning corrects the ESCS of heavy ions and the simulated primary radiation results, which will be discussed in section 3.2.

σ_e^{ML} of different ions into Si target and Au ions in different targets are compared with experiments as shown in Fig. 7. σ_e^{ML} show excellent agreement with the experiment points near and above the Bragg Peaks, both for H, Li, B, C, O, Si, Kr and Au (bottom to top) ions in Si and for Au ions into C, Si, Al, Cu and Au (bottom to top) targets, as shown in Fig. 7 (a) and (b), respectively. Considering the difficulty in obtaining the ESCS in both theory and experiments, SRIM, BT and UCA are accurate only for specific systems, particularly in low energy ranges. The machine learning can overcome the difficulties, with mining the complex relation between the data selected from the three theoretical databases and obtain an optimized database compare to the three theoretical models.

In addition, UCA is known as performing well for gas and limiting for metals so the effect of UCA to the machine learning model is tested at low energies. The ESCS trained without UCA dataset for U ions into Au is shown in Fig. 6(f). The deviation between the machine learning results trained with and without UCA is 54.6 %, lower than the systematic error of 94.6 % between the training dataset for energies lower than 100 keV/u. For energies higher than 100 keV/u, their deviation is negligible. Moreover, by training different systems of heavy and light ions into gases and metals with and without UCA, we calculated the deviation of the training results of the two cases with experiments in IAEA database, respectively. Their difference is within 5 % which can be ignored.

3.2. *Verification of ion range*

The accuracy of σ_e^{ML} is further verified by comparing IM3D simulated ion depth distribution with the experimental results. For 400 keV He ions implanted into W target, the ion depth distributions simulated by IM3D with σ_e^{SRIM}, σ_e^{BT}, σ_e^{UCA} and σ_e^{ML} are shown in Fig. 8 (a) [36]. It can be concluded that the ion depth distributions simulated with σ^{SRIM} and σ_e^{ML} show better agreement with the experiments compared to those with other ESCS. BT and UCA underestimate and overestimate the ion depth distribution, respectively, due to their inaccurate calculation of ESCS at energy range lower than 10 keV/u, which can be seen in Fig. 6(a).

Fig. 8(b) shows the ion depth distribution of 80 keV/u C ion into Si target. The ion depth distribution simulated with σ_e^{SRIM} and σ_e^{BT} show the most agreement with the experimental results [37]. For the result simulated with σ_e^{ML}, it shows a slight deviation of 4.65 % near the peak position of the depth distribution. The result simulated with σ_e^{UCA} shows an obvious overestimation. Fig. 8(c) shows the ion depth distributions of

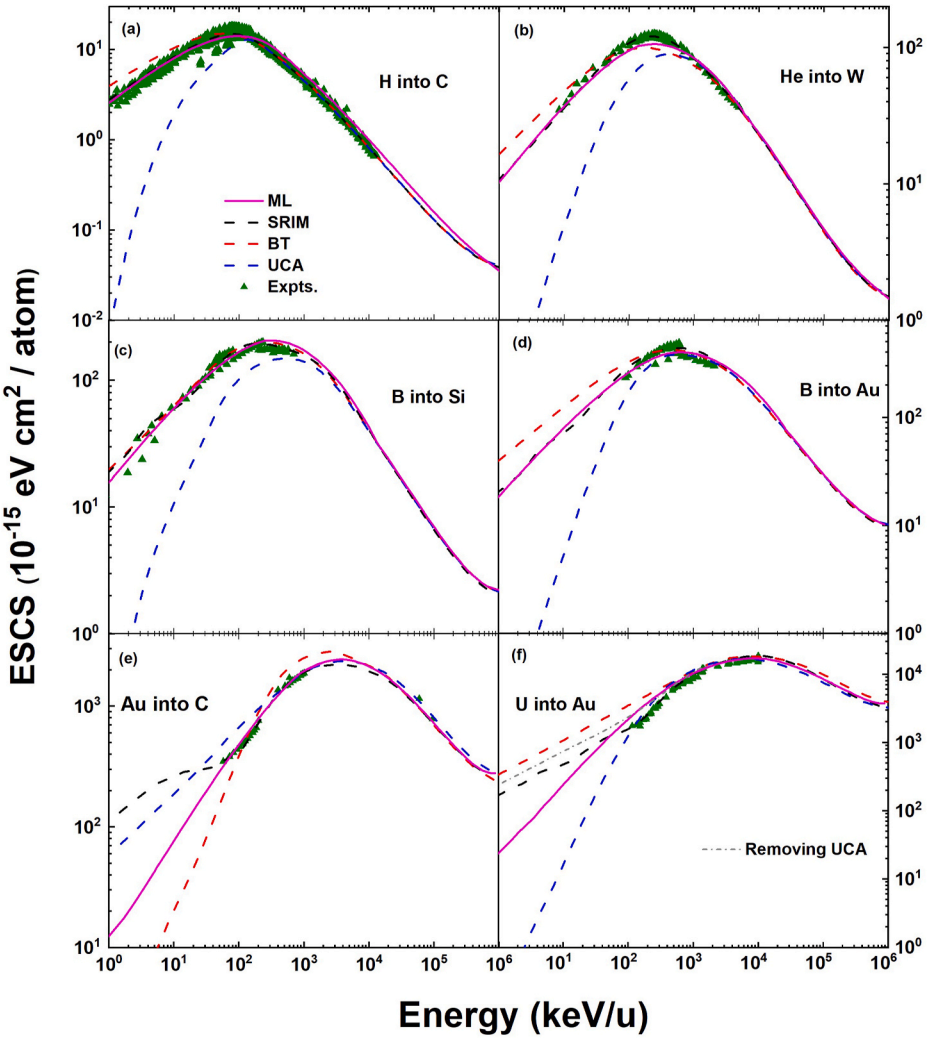

Fig. 6. Comparison of σ_e^{ML}, σ_e^{SRIM}, σ_e^{BT}, σ_e^{UCA} and experimental data [25] for (a) H ions into C, (b) He ions into W, (c) B ions into Si, (d) B ions into Au, (e) Au ions into C and (f) U ions into Au. For U ions into Au, σ_e^{ML} without UCA dataset in the low energy range has also been given as gray dash-dot line for comparison.

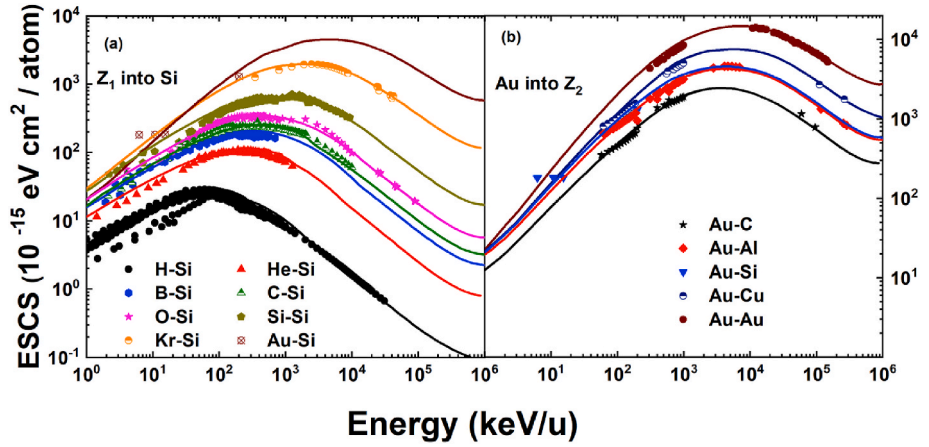

Fig. 7. Comparison of σ_e^{ML} with experimental data [25] of (a) different ions into Si and (b) Au ions into different targets.

2 MeV Au ions into Si target [6]. The results simulated with σ_e^{ML} and σ_e^{UCA} have a good prediction to the ion range for heavy ion irradiation, while results with σ_e^{ML} and both underestimate the ion range. The effect of including UCA in the training dataset on the ion range is also tested. By simulating the ion radiation with the machine learning results with and without UCA shown in Fig. 6(f) at energies of 10 keV/u, 100 keV/u and 1 MeV/u, the average deviation of their ion range results is only within 7 %.

IM3D simulated ion ranges with different ESCS are also compared at different incident energies for typical heavy (take Au ion as an example)

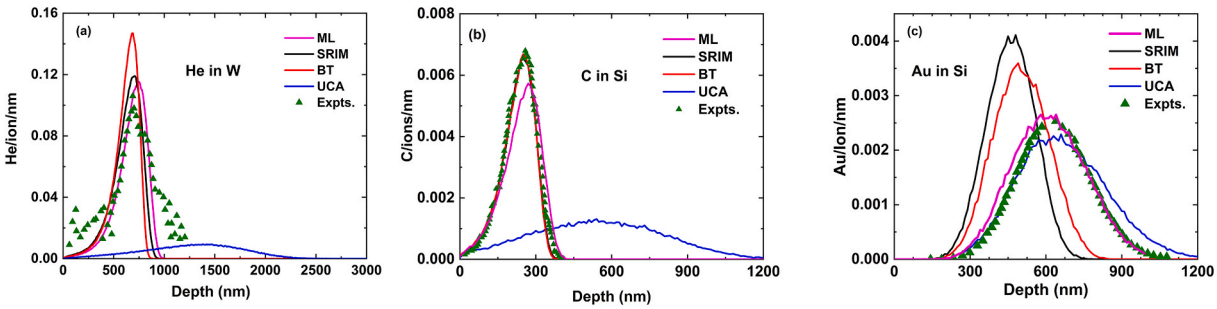

Fig. 8. Ion depth distributions of experimental results and IM3D simulated with σ_e^{SRIM}, σ_e^{BT}, σ_e^{UCA} and σ_e^{ML} of (a) He ions into W [36], (b) C ions into Si [37] and (c) Au ions into Si [6].

and light ions (take B ions as an example). For B ions in Fig. 9(a), the ion range on a wide energy range (30–300 keV) simulated with σ_e^{ML} are consistent with the experiments [38] with the same accuracy of σ_e^{SRIM} and σ_e^{BT}. Results simulated with σ_e^{UCA} overestimate the ion range. For Au ions, the ion ranges simulated with σ_e^{UCA} and σ_e^{ML} are more consistent with the experiments [39], while those simulated with σ_e^{SRIM} and σ_e^{BT} have an underestimation values. Figs. 8 and 9 show that σ_e^{ML} can give the accurate ion depth distribution and ion range for light and heavy ions.

The ion ranges simulated with σ_e^{SRIM}, σ_e^{BT}, σ_e^{UCA} and σ_e^{ML} are generally compared with each other for arbitrary systems and incident energies. The projectile ions and targets are both randomly selected in elements from H to U, and the incident ion energies are randomly selected in the energy range of 0 – 10 keV/u, 0 – 100 keV/u and 0 – 1000 keV/u, respectively, in which 1000 times IM3D simulation are performed with each ESCS to obtain the average ion ranges. In order to be in consist with the training dataset, the average ion ranges simulated with σ_e^{SRIM}, σ_e^{BT} and σ_e^{UCA} are taking as the reference values. The errors between the reference values and the results simulated with the four different ESCS are show in Fig. 10. The machine learning performs a neutralization to a certain extent on the three theoretical ESCS. In Fig. 6, it shows that σ_e^{ML} are in the middle of σ_e^{SRIM}, σ_e^{BT} and σ_e^{UCA} at low energies. Therefore, the error with σ_e^{ML} is always the smallest one than that with the theoretical data.

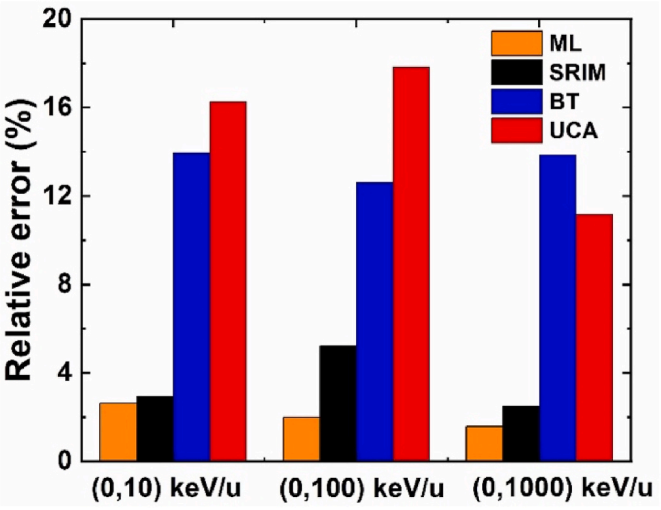

Fig. 10. Relative error between the average ranges simulated by IM3D with σ_e^{SRIM}, σ_e^{BT}, σ_e^{UCA}, σ_e^{ML} and the reference values (average ranges simulated with σ_e^{SRIM}, σ_e^{BT} and σ_e^{UCA}).

3.3. Analysis of reciprocity

The average relative coherent coefficient (γ_{rec}) of σ_e^{ML}, σ_e^{SRIM}, σ_e^{BT} and σ_e^{UCA} for ion-target combinations with atomic numbers from 1 to 92 are calculated in the energy range of 1–25 keV/u. In Fig. 11, the blue zone indicates the high consistence of the Z_1-Z_2 pair, while the yellow and red zone show the invalid reciprocity with γ_{rec} reaches 0.8–1.0. The lower the coefficient, the better in validation. For SRIM in Fig. 11(b), γ_{rec} in the systems of heavy ions and heavy targets reach 0.8. γ_{rec} of σ_e^{UCA} are 0.8 for most light and heavy ions. The calculated denominator in Eq. (5) is small leading to the higher coefficients because σ_e^{UCA} at lower energy (below several keV/u) is small. σ_e^{ML} and σ_e^{BT} are the best consistence with reciprocity as they show almost blue zone for all the systems, especially for

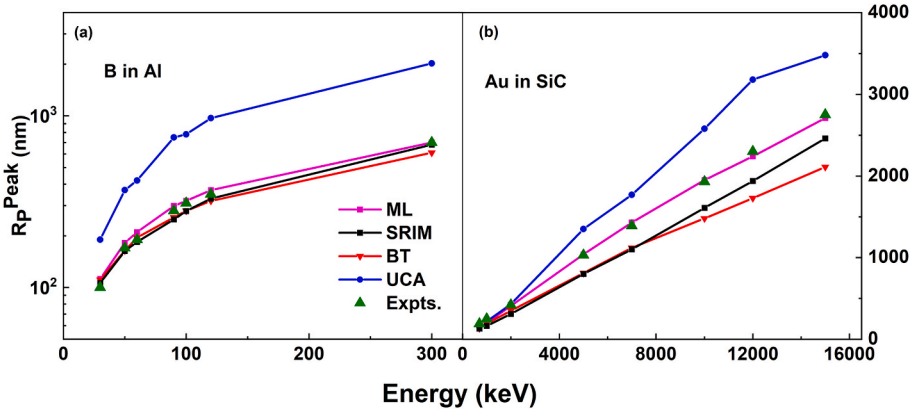

Fig. 9. Comparison between simulated and experimental peak positions with σ_e^{SRIM}, σ_e^{BT}, σ_e^{UCA} and σ_e^{ML} of (a) B ions into Al [38] and (b) Au ions into SiC with different incident energies [39].

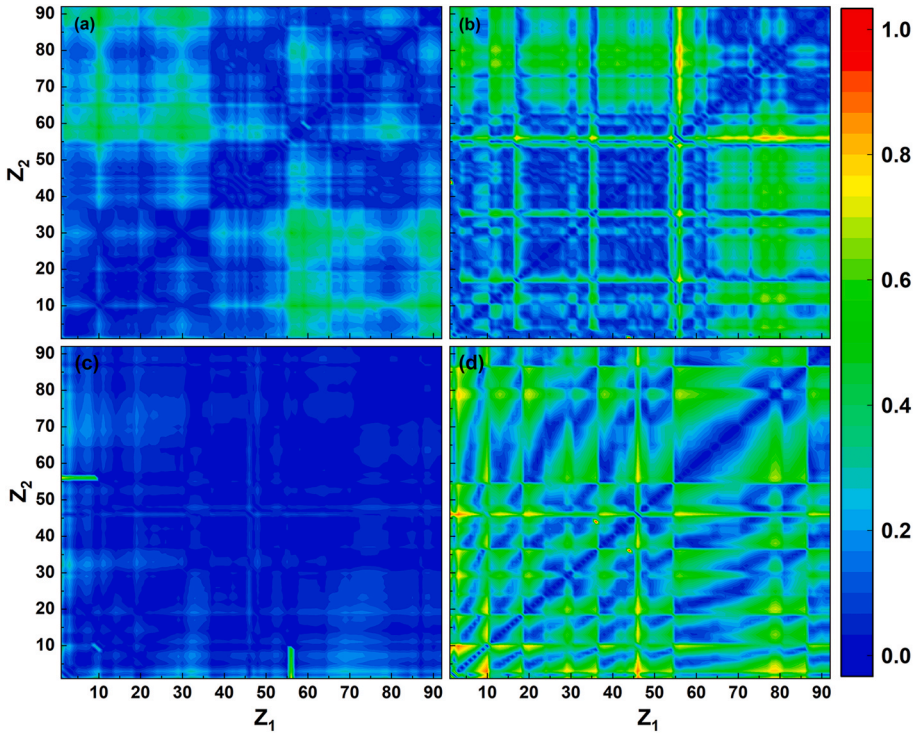

Fig. 11. Average relative coherent coefficient (γ_{rec}) calculated using (a) (a) σ_e^{ML} (b) σ_e^{SRIM} (c) σ_e^{BT} and (d) σ_e^{UCA}.

BT. However, the better γ_{rec} of σ_e^{BT} does not represent its high accuracy for its overestimation for heavy ions and thus underestimates ion depth distribution, as shown in Fig. 8(a) and (c) γ_{rec} of σ_e^{ML} is slightly higher for the heavy ions and target systems and reaches about 0.4. Such low value does not affect the accuracy of σ_e^{ML} as shown in Fig. 6(e) and (f) as well as the ion depth distribution simulations in Figs. 8(c) and 9(b). In fact, the reciprocity theory is not directly related to the accuracy of the electronic stopping. After the validation with the insufficient experimental ESCS and ion ranges, the machine learning have been validated to be more accurate than SRIM, BT and UCA. Here the reciprocity theory is applied to validate the reasonability of all the ion-target combinations in which the experiments data are absent.

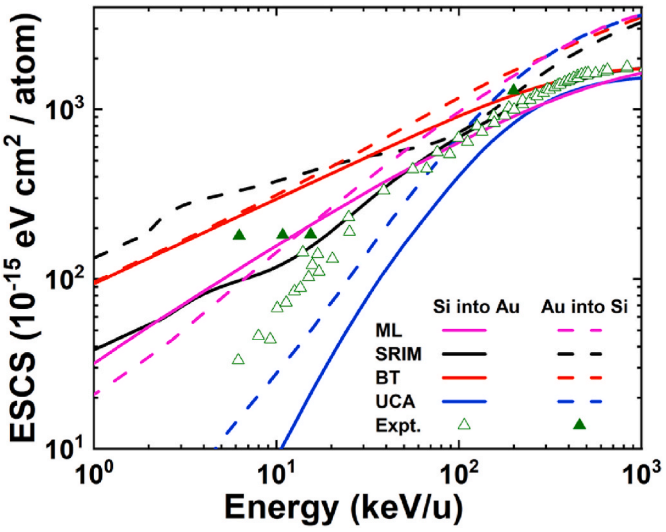

Fig. 12. Comparison of σ_e^{ML}, σ_e^{SRIM}, σ_e^{BT} and σ_e^{UCA} with the experimental values for Si ions into Au and Au ions into Si systems.

Fig. 12 shows the theoretical and experimental ESCS of Au ions into Si and Si ions into Au targets. σ_e^{ML} shows a good consistent with the experimental values. At energies lower than 30 keV/u, σ_e^{ML} of Au ions into Si shows good agreement with σ_e^{ML} of Si ions into Au. SRIM agrees with the experimental value for Si ions into Au with the energies higher than 10 keV/u, but it overestimates the ESCS of Au ions into Si. BT and UCA overestimate and underestimate the ESCS of the two systems, respectively. Based on the overall reciprocity analysis and the comparison of Au ions into Si and Si ions into Au systems, it can be concluded that σ_e^{ML} shows good reciprocity performance and is consistent with experimental values.

3.4. Extracting key descriptors with machine learning

The good accuracy of σ_e^{ML} originates from the reasonably selected training dataset and descriptors. Particularly, LASSO excels in extracting the most critical descriptors, which are obtained based on the existing theories. In addition, the exponent of each element in the theories are extended to a wider range, enabling a more comprehensive exploration within the training data. Here the descriptors importance I_m of the m-th descriptor for a certain material is defined according to Ref. [24] as $I_m = |a_m F_m| / \sum_{i=1}^{n} |a_i F_i|$, where a is the coefficient of descriptors, F is the descriptor value, and n is the total number of descriptors.

The top-ten important descriptors of low- and high-energy ranges are shown in Fig. 13. The upper shadowed parts in the columns on the right of each panel summarize the importance of descriptors with minor importance. The upper red parts represent the importance increase when elements are accumulated following the importance order from high to low. In the low energy range, the top ten important descriptors account for 39.80 %. Most of them are elements in Eqs. (1) and (2). In the descriptor pool, there are 384 descriptors, each in the same form as Eqs. (1) and (2). With changing the exponent of each term, the sum importance of the descriptors with the same form as Eq. (1) and Eq. (2) are 10.98 % and 21.89 %, respectively. Since the difference between Eq. (1) and Eq. (2) is that, Z_1 and Z_2 is replaced by the effective value of Z_{1S}

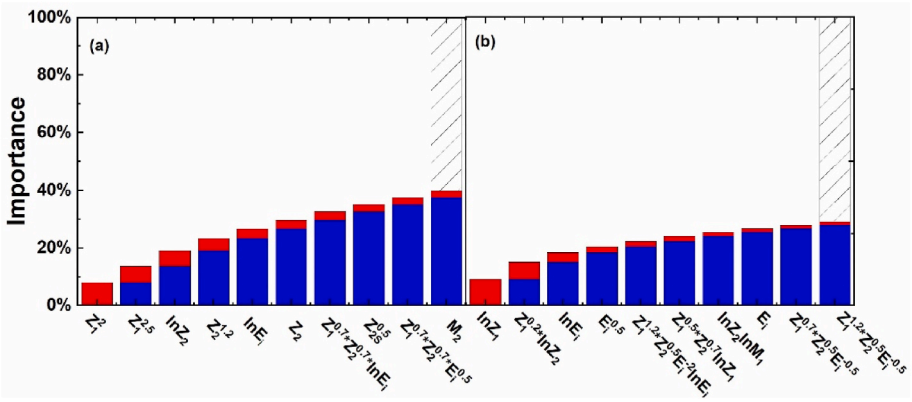

Fig. 13. Percentage histogram of the importance ratio for LASSO-selected descriptors for a) low and b) high energies.

and Z_{2S} adjusted by the term in the Hartree-Fock-Slater screening function. It shows the importance of correcting the ion effective charge. Besides, the number of descriptors for shell and barkas correction is 96, with the sum importance of 19.96 %. In the high energy range, the sum of the top-ten descriptors is only 29.04 %. For the key descriptors shown in Fig. 13(b), $Z_1^{1.2}Z_2^{0.5}E_i^{-2}\ln E_i$, $Z_1^{0.7}Z_2^{0.5}E_i^{-0.5}$ and $Z_1^{1.2}Z_2^{0.5}E_i^{-0.5}$ have the same form as the Bethe-Bloch and Bohr theories of Eq. (3), with the total importance of 4.21 %. The other key descriptors are the elements in the Bethe-Bloch or Bohr theories.

In both low and high energies, the key descriptors maintain the similar form to the reference formula of the LSS and BBB theories, respectively, with adjusting the exponents of each element. The adjusted exponents of the elements are set in the descriptor pool as discussed in Section 2.2. Machine learning can optimize the formulas based on the training data. These key descriptors ensure the accuracy of the predicted ESCS and simulated primary radiation damage, providing valuable insights for future ESCS modeling.

4. Conclusion

Based on the experimental-cleaned theoretical databases of SRIM, BT and UCA, and the descriptor pool extracted from existing theoretical models, the machine learning study of the universal ESCS database of ions in matter is performed with the LASSO algorithm. The database contains the data of ion-target combinations with atomic numbers from 1 to 92 in a wide energy range of 1 keV/u - 1 GeV/u. By comparing with experiments and IM3D simulations, the machine learning predicted ESCS shows a overall higher accurate prediction of ESCS and ion depth distribution/range compared to that with SRIM, BT and UCA for both light and heavy ions. Moreover, σ_e^{ML} has a good reciprocity performance with the analysis of the average relative coherent coefficient. The accuracy of σ_e^{ML} comes from the key descriptors selected by LASSO, which maintain the similar form to the reference formula of the LSS and B-B-B theories, with adjusting the exponent of each element. Thus, machine learning could mine the complex relation between the training data to achieve data optimization with the key descriptors. Finally, a universal machine learning ESCS database (ESDB) is established and available on the website http://theory.issp.ac.cn/dbpro/. The database provides an accurate description of the ion energy loss in matter, which is significant for the simulation of primary damage, which has vast applications in fields like semiconductor manufacturing and nano-structures fabrication. This work didn't consider the physical state and charge state directly, which may influence the electronic stopping more or less. Further works including these effects are still needed to improve the machine learning performance.

CRediT authorship contribution statement

Fan Cheng: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xun Liu:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis. **Qirong Zheng:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Chuanguo Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis. **Bo Da:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis. **Yonggang Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 12375277), the Anhui Provincial Natural Science Foundation (Grant No. 2308085J04) and the outstanding member of Youth Innovation Promotion Association CAS (Grant No. Y202087). Some of the calculations were performed at the Center for Computational Science of CASHIPS, the ScGrid of Supercomputing Center, and the Computer Network Information Center of the Chinese Academy of Sciences.

References

[1] S. Lohmann, D. Primetzhofer, Disparate energy scaling of trajectory-dependent electronic excitations for slow protons and He ions, Phys. Rev. Lett. 124 (9) (2020), https://doi.org/10.1103/PhysRevLett.124.096601.
[2] S. Coloma, et al., The effect of ionizing radiation on robotic trajectory movement and electronic components, Nucl. Eng. Technol. 55 (11) (2023) 4191–4203, https://doi.org/10.1016/j.net.2023.07.041.
[3] C.G. Zhang, et al., Fast generation of reliable primary radiation damage of BCC tungsten by sampling molecular dynamics databases, Nucl. Mater. Energy 35 (2023) 101443, https://doi.org/10.1016/j.nme.2023.101443.
[4] G.S. Was, Fundamentals of Radiation Materials Science, Springer, Berlin Heidelberg New York, 2007, p. 827.

[5] F. Akbari, et al., Predicting electronic stopping powers using stacking ensemble machine learning method, Nucl. Instrum. Methods Phys. Res., Sect. B 538 (2023) 8–16, https://doi.org/10.1016/j.nimb.2023.02.023.
[6] Y.Q. Chang, et al., MeV Au ion irradiation in silicon and nanocrystalline zirconia film deposited on silicon substrate, Nucl. Instrum. Methods Phys. Res., Sect. B 286 (2012) 173–179, https://doi.org/10.1016/j.nimb.2012.01.017.
[7] F. Cheng, et al., Sensitivity of ion implantation to low-energy electronic stopping cross-sections, Radiat. Phys. Chem. 204 (2023), https://doi.org/10.1016/j.radphyschem.2022.110681.
[8] X. Guo, et al., Development of an electronic stopping power model based on deep learning and its application in ion range prediction, Chin. Phys. B 31 (7) (2022), https://doi.org/10.1088/1674-1056/ac4e0c.
[9] N. Bohr, On the theory of the decrease of velocity of moving electrified particles on passing through matter, Philos. Mag. A 25 (145) (1913) 22, https://doi.org/10.1080/14786440108634305.
[10] H. Bethe, On the theory of the passage of rapid charged particle radiation through matter, Ann. Phys. (Berlin, Ger.) 5 (3) (1930) 76.
[11] F. Bloch, The slow down of rapidly moving particles in the their passing through solid matter, Ann. Phys. (Berlin, Ger.) 16 (3) (1933) 36.
[12] N.R. Arista, A.F. Agustín, Non-linear approach to the energy loss of ions in solids, Adv. Quantum Chem. 45 (2004) 31, https://doi.org/10.1016/s0065-3276(04)45003-5.
[13] J. Lindhard, M. Scharff, Energy disspation by ions in KeV region, Phys. Rev. 124 (1) (1961) 128, https://doi.org/10.1103/PhysRev.124.128.
[14] J.F. Ziegler, M.D. Ziegler, J.P. Biersack, Srim – the stopping and range of ions in matter (2010), Nucl. Instrum. Methods Phys. Res., Sect. B 268 (2010) 6, https://doi.org/10.1016/j.nimb.2010.02.091.
[15] P. Sigmund, A. Schinner, Binary stopping theory for swift heavy ions, Eur. Phys. J. D 12 (2000) 10, https://doi.org/10.1140/epjd/e2005-00323-2.
[16] P.L. Grande, G. Schiwietz, Impact-parameter dependence of the electronic energy loss of fast ions, Phys. Rev.A 58 (1998), https://doi.org/10.1016/j.nimb.2004.12.010.
[17] E.D. Cantero, et al., Experimental and theoretical study of the energy loss of C and O in Zn, Phys. Rev.A 84 (1) (2011), https://doi.org/10.1103/PhysRevA.84.014902.
[18] H. Paul, A. Schinner, Empirical stopping power tables for ions from 3Li to 18Ar and from 0.001 to 1000MeV/nucleon in solids and gases, Atomic Data Nucl. Data Tables 85 (2) (2003) 377–452, https://doi.org/10.1016/j.adt.2003.08.003.
[19] K. Wittmaack, Misconceptions impairing the validity of the stopping power tables in the SRIM library and suggestions for doing better in the future, Nucl. Instrum. Methods Phys. Res., Sect. B 380 (2016) 57–70, https://doi.org/10.1016/j.nimb.2016.04.057.
[20] A.J. Jinia, et al., Intelligent Radiation: a review of Machine learning applications in nuclear and radiological sciences, Ann. Nucl. Energy 201 (2024) 110444, https://doi.org/10.1016/j.anucene.2024.110444.
[21] W.A. Parfitt, R.B. Jackman, Machine learning for the prediction of stopping powers, Nucl. Instrum. Methods Phys. Res., Sect. B 478 (2020) 21–33, https://doi.org/10.1016/j.nimb.2020.05.015.
[22] Mehnaz, et al., Ensemble machine learning methods: predicting electron stopping powers from a small experimental database, Phys. Chem. Chem. Phys. 23 (10) (2021) 6062–6074, https://doi.org/10.1039/d0cp06521h.
[23] F. Bivort Haiek, et al., ESPNN: a novel electronic stopping power neural-network code built on the IAEA stopping power database. I. Atomic targets, J. Appl. Phys. 132 (24) (2022), https://doi.org/10.1063/5.0130875.
[24] X. Liu, et al., Unveiling the principle descriptor for predicting the electron inelastic mean free path based on a machine learning framework, Sci. Technol. Adv. Mater. 20 (1) (2019) 1090–1102, https://doi.org/10.1080/14686996.2019.1689785.
[25] For "electronic stopping power of matter for ions", 2020; Available from: https://www-nds.iaea.org/stopping, 2020.
[26] C.C. Montanari, et al., The IAEA electronic stopping power database: modernization, review, and analysis of the existing experimental data, Nucl. Instrum. Methods Phys. Res., Sect. B 551 (2024) 165336, https://doi.org/10.1016/j.nimb.2024.165336.
[27] Y.G. Li, et al., IM3D: a parallel Monte Carlo code for efficient simulations of primary radiation displacements and damage in 3D geometry, Sci. Rep. 5 (2015) 18130, https://doi.org/10.1038/srep18130.
[28] P. Sigmund, Reciprocity in the electronic stopping of slow ions in matter, Eur. Phys. J. D 47 (1) (2008) 45–54, https://doi.org/10.1140/epjd/e2008-00011-9.
[29] Pedregosa, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 6, https://doi.org/10.5555/1953048.2078195.
[30] R. Tibshirani, Regression shrinkage and selection via the Lasso, J.R.Stat.Soc.B 58 (1) (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[31] H. Sugiyama, Modication of lindhard-scharff-schiott formula of electronic stopping power, J. Phys. Soc. Jpn. 50 (3) (1980) 929–932, https://doi.org/10.1143/JPSJ.50.929.
[32] Neetu, et al., Electronic stopping power of polymers for heavy ions in the ion energy domain of LSS theory, Radiat. Meas. 44 (4) (2009) 363–368, https://doi.org/10.1016/j.radmeas.2009.03.031.
[33] J.E. Asri, et al., Electronic stopping powers of formvar and mylar polymeric materials for heavy ions: LSS modified theory and Monte Carlo simulation, Nucl. Technol. 205 (9) (2019) 1236–1244, https://doi.org/10.1080/00295450.2019.1590071.
[34] P. Sigmund, Particle penetration and radiation effects, in: Penetration of Atomic and Molecular Ions, vol. 2, Springer International Publishing Switzerland, 2006, p. 617.
[35] F. Cheng, et al., Electronic energy loss assessment in theoretical modeling of primary radiation damage in tungsten, Int. J. Mod. Phys. C 32 (2021) 15, https://doi.org/10.1142/S0129183121501345.
[36] L. Qiao, et al., Helium retention depth profile in plasma-facing materials measured by glow-discharge optical emission spectroscopy, Spectrochim. Acta, Part B 182 (2021) 106257, https://doi.org/10.1016/j.sab.2021.106257.
[37] S. Intarasiri, et al., RBS and ERDA determinations of depth distributions of high-dose carbon ions implanted in silicon for silicon–carbide synthesis study, Nucl. Instrum. Methods Phys. Res., Sect. B 249 (1–2) (2006) 859–864, https://doi.org/10.1016/j.nimb.2006.03.182.
[38] D. Fink, et al., 30 keV to 2 MeV Boron implantation profiles in solids, Radiat. Eff. Defects Solids 115 (1–3) (2006) 93–112, https://doi.org/10.1080/10420159008220558.
[39] K. Jin, et al., Ion distribution and electronic stopping power for Au ions in silicon carbide, Nucl. Instrum. Methods Phys. Res., Sect. B 307 (2013) 65–70, https://doi.org/10.1016/j.nimb.2013.02.051.