Machine learning approach for the prediction of electron inelastic mean free paths

Xun Liu,[1,2] Lihao Yang,[1,2] Zhufeng Hou,[3] Bo Da,[2,4,*] Kenji Nagata,[2] Hideki Yoshikawa,[2] Shigeo Tanuma,[2] Yang Sun,[5] and Zejun Ding[1,†]

[1]*Hefei National Laboratory for Physical Sciences at Microscale and Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*

[2]*Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan*

[3]*State Key Laboratory of Structural Chemistry, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou 350002, China*

[4]*Research Center for Advanced Measurement and Characterization, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*

[5]*Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, 10027, USA*

(Received 28 May 2020; accepted 12 February 2021; published 24 March 2021)

The prediction of electron inelastic mean free paths (IMFPs) from simple material parameters is a challenging problem in studies using electron spectroscopy and microscopy. Herein, we propose a machine learning (ML) approach to predict IMFPs from some basic material property data. The ML model showed excellent performance based on the calculated IMFPs for a group of 41 elemental materials [Li, Be, C (graphite), C (diamond), C (glassy), Na, Mg, Al, Si, K, Sc, Ti, V, Cr, Fe, Co, Ni, Cu, Ge, Y, Nb, Mo, Ru, Rh, Pd, Ag, In, Sn, Cs, Gd, Tb, Dy, Hf, Ta, W, Re, Os, Ir, Pt, Au, and Bi] from a previous paper [Shinotsuka *et al.*, Surf. Interface Anal. **47**, 871 (2015); **47**, 1132 (2015)]. which was comparable with that of the robust Tanuma-Powell-Penn (TPP-2M) formula. The developed ML model was then extended to materials that do not have reported IMFPs in the database. The IMFPs for 18 transition and lanthanide metals (Mn, Zn, Zr, Tc, Cd, La, Ce, Pr, Nd, Pm, Sm, Eu, Ho, Er, Tm, Yb, Lu, and Hg) were predicted by the ML model. In the comparison with full-Penn algorithm-calculated IMFPs through two newly found experimental energy loss functions (ELFs), i.e., Mn and Zr, the Gaussian process regression-predicted IMFPs not only agreed well with those calculated using the TPP-2M formula in the energy range >50 eV but were also consistent with the trend of IMFPs calculated through experimental ELFs in the range of 2.7–50 eV, where the TPP-2M formula cannot be used. Our findings suggest that ML is very powerful and efficient and has great potential to complete a database of IMFPs for materials that can prove solutions closer to reality than empirical models on materials with similar physical and chemical properties and can be applied to other different situations for correlated information prediction.

DOI: 10.1103/PhysRevMaterials.5.033802

I. INTRODUCTION

The inelastic mean free path (IMFP), which describes the mean distance an electron travels through a solid before losing energy [1], is an essential parameter in determinations of surface sensitivity for surface electron spectroscopies, such as x-ray photoelectron spectroscopy (XPS) [2–4], Auger electron spectroscopy (AES) [2,5–7], reflection electron energy loss spectroscopy (REELS) [8,9], and for quantitative analyses with these techniques. Additionally, IMFP is one of the most important constants in the Monte Carlo method [10], and some other important parameters can be successively obtained by the simulation of the physical process of the incident-electron scattering in materials to observe electron transport behaviors, represented by theoretically determined parameters of materials, including the mean escape depth [11], backscattering factor [12,13], surface excitation parameters [14–16], and so on. Since the IMFP is fundamentally important for both experimental and theoretical studies, several methods, e.g., full-Penn algorithm (FPA) [17], Mermin algorithm [18], and ex-Mermin algorithm [19], have been established to calculate IMFPs at electron energies >50 eV. As the most popular and reliable theoretical algorithms for the calculation of IMFPs, the difference between them is the usage of the dielectric function. Firstly, the most well-accepted algorithm is the FPA, which was proposed by Penn [17]. In the calculation of FPA, Penn used the Lindhard dielectric function to represent the probability for inelastic scattering; however, the finite lifetime broadening of the plasmon is neglected. Tanuma *et al.* frequently calculated IMFPs with FPA: 27 elemental materials [20,21], 15 inorganic compounds [22], and 14 organic compounds [23] in the energy range from 50 to 2000 eV; and Shinotsuka *et al.* calculated 41 elemental materials [24] and 42 inorganic compounds [25] in the energy range from 50 to 200 keV. Secondly, for higher accuracy, Mermin developed the Mermin model [18], which considered the finite

*da.bo@nims.go.jp
†zjding@ustc.edu.cn

lifetime broadening of the plasmon, using the so-called Mermin dielectric function in IMFPs calculation. Abril *et al.* also brought out a series of IMFPs calculations [26] with the Mermin method. Recently, the latest brand-new extended Mermin method [19] was proposed by Da *et al.*; they further improve the calculation accuracy of the Mermin method. However, all these methods rely on accurate determination of the energy loss function (ELF) because they are built in the framework of the dielectric theory. Therefore, these theoretical approaches can only be applied to materials that have experimental optical constants available to provide the ELFs for fitting [27]. Unfortunately, until now, IMFPs for numerous materials, even for elemental materials, still cannot be calculated through these algorithms due to the lack of reliable experimental data for their optical constants over a sufficiently wide energy range. Specifically, the preparation of a lanthanide metal sample with a sufficiently clean surface requires quite a high level of facilities and skills. In detail, in the polishing process of lanthanide metal sample, the surface is very susceptible to contamination. However, in recent years, many groups including our group successfully obtained ELFs from experimental REELS spectra [28–32] and successfully obtained ELFs of multiple materials [33–37], except for lanthanide metals. Until now, there are still numerous materials that lack reliable experimental data for their optical constants over a sufficiently wide energy range because of the lack of sufficiently accurate REELS spectra. Therefore, it is critical to develop an alternative way to determine the unknown IMFPs based on lessons learned from those materials with well-established optical data.

Recently, several methods have been proposed to characterize parameters for electron scattering, which is the history of effective attenuation lengths (EALs) and IMFPs, based on those well-established material features. Details of some of the well-known empirical formulas for them are provided as follows.

(i) The universal formula is

$$\lambda_{\mathrm{EAL}} = AE^{-2} + BE^{1/2}, \tag{1}$$

where A and B are coefficients that are fitted according to the kinds of material being investigated, including element, inorganic compounds, and organic compounds.

Our starting point for the development of a predictive EAL formula is the paper by Seah and Dench [38]. It must be stated that, at the time that the review was published, it was believed that IMFPs and attenuation lengths (now known as EALs) were the same quantity. More detailed information of IMFPs and EALs can be found in Refs. [39,40]. The universal formula developed in the Seah and Dench review [38] was based on an analysis of what are now known as EALs, not IMFPs. Their universal formula results from a least squares analysis of the attenuation lengths in solids for energies $<10\,000$ eV above the Fermi level. They considered solid materials including elements, inorganic compounds, organic compounds, and gas adsorbates. Their paper initiated an idea in attenuation lengths description, namely, to seek a universal approach to describe the attenuation lengths by a simple formula suitable for all materials. However, the individual EAL measurements had large uncertainties [40].

(ii) The Bethe equation is

$$\lambda = \frac{E}{E_p^2[\beta \ln(\gamma E)]}, \tag{2}$$

where E_p is the free-electron plasmon energy, and β and γ are coefficients.

A basic equation to calculate IMFPs is the Bethe equation, which is the starting point of the robust Tanuma-Powell-Penn (TPP-2M) equation [24]. The parameters used in the original Bethe equation [41] are microscopic quantities. Tanuma *et al.* [21] transformed the parameters used in the Bethe equation into macroscopic quantities, in other words, material-dependent parameters, and expanded the Bethe equation to low energies (<200 eV), thus making it an empirical formula. However, the Bethe equation is only valid for energies >200 eV. Thus, the TPP-2M equation was developed [23] based on the Bethe equation and later expressed for relativistic electron energies [24].

(iii) The TPP-2M equation [24] is

$$\begin{aligned}
\lambda &= \frac{\alpha(E)E}{E_p^2\left\{\beta_r \ln[\gamma_r \alpha(E)E] - \frac{C_r}{E} + \frac{D_r}{E^2}\right\}}\,(\mathrm{nm}),\\
\alpha(E) &= \left[1 + \left(\frac{E}{2m_e c^2}\right)\right] \Big/ \left[1 + \left(\frac{E}{m_e c^2}\right)\right]^2,\\
\beta_r &= -1.0 + \frac{9.44}{\left(E_p^2 + E_g^2\right)^{0.5}} + 0.69\rho^{0.1}\,(\mathrm{eV^{-1}nm^{-1}}),\\
\gamma_r &= 0.191\rho^{-0.5}\,(\mathrm{eV^{-1}}),\\
C_r &= 19.7 - 9.1U\,(\mathrm{nm^{-1}}),\\
D_r &= 534 - 208U\,(\mathrm{eVnm^{-1}}),\\
U &= \frac{N_v \rho}{M} = \left(\frac{E_p}{28.816}\right)^2,
\end{aligned} \tag{3}$$

where $m_e c^2$ is the electron rest energy (510 998.9 eV), E_p is the free-electron plasmon energy (in electronvolts), E_g is the bandgap energy for nonconductors (in electronvolts), ρ is the bulk density (in grams per cubic centimeter), and N_v is the number of valence electrons per atom or molecule.

(iv) The TPP-LASSO-S formula is [42]

$$\begin{aligned}
\lambda &= \frac{\alpha(E)E}{E_p^2\{\beta_r \ln[\gamma_r \alpha(E)E]\}}\,(\text{Å}),\\
\alpha(E) &= \left[1 + \left(\frac{E}{2m_e c^2}\right)\right] \Big/ \left[1 + \left(\frac{E}{m_e c^2}\right)\right]^2,\\
\beta_r &= -0.0012 + 0.046\left(\frac{M}{\rho N_v}\right)^{0.5} - 0.035\left(\frac{M}{\rho N_v}\right)^{0.4}\\
&\quad + 0.0019\frac{Z}{N_v}\,(\mathrm{eV^{-1}Å^{-1}}),\\
\gamma_r &= -0.07 + 0.26[\rho(E_i + E_g)]^{-0.2}\\
&\quad + 0.066\left(\frac{Z\rho}{M}\right)^{-0.8}\,(\mathrm{eV^{-1}}),
\end{aligned} \tag{4}$$

where $m_e c^2$ is the electron rest energy (510 998.9 eV), E_p is the free-electron plasmon energy (in electronvolts), E_i is the starting-point energy (in electronvolts), E_g is the bandgap

energy for nonconductors (in electronvolts), ρ is the bulk density (in grams per cubic centimeter) and N_v is the number of valence electrons per atom or molecule.

Additionally, an empirical formula that includes an energy region <50 eV is also proposed in Ref. [43]. Although the formula can describe electron IMFPs of a low energy region (<50 eV), according to Ref. [43], the fitting coefficients cannot be related to material-dependent parameters. In other words, for now, the formula needs fitting according to IMFP curves for every certain material. The formula is different for each material, not to mention the predictive power it will have.

Recently, Shinotsuka *et al.* [24] developed a relativistic version of the TPP-2M equation based on the modified Bethe equation by introducing a data-driven concept. In the equation, the formulas for β and γ were confirmed in Ref. [23]. Through application of the TPP-2M equation to a variety of materials, the two correction terms C and D were also introduced into the denominator to adjust the prediction results for energies <200 eV in Ref. [23]. Lastly, a relativistic modification was added, which gave β_r, γ_r, C_r, D_r, and the modification term $\alpha(E)$, to provide the complete form of the TPP-2M equation, thus representing another step toward an accurate description of electron IMFPs using material-dependent parameters. Using the TPP-2M equation, IMFPs can be estimated for any material.

The TPP-2M formula [24], which is a modified form of the Bethe equation [41], shows a robust fitting of calculated IMFPs and can predict IMFPs using material properties, such as density and the number of electrons per atom, in the electron energy range >50 eV. Tanuma *et al.* have worked on the TPP-2M formula from the initial paper [20] to the most recent by Shinotsuka *et al.* [24] and expanded their original IMFP calculations (additional materials, expanded energy range, and improved calculation accuracy). In the development process of the TPP-2M, Tanuma *et al.* were using FPA-calculated IMFP data as a target. However, the FPA-calculated IMFPs have a relatively poorer accuracy for energies <50 eV and >200 keV due to the uncertainty of the exchange-correlation potential [21] and the neglect of the transverse differential cross-section for inelastic scattering [24], respectively. Although the main motivation for the IMFP calculations of Tanuma *et al.* was to provide needed IMFP data for practical applications of AES and XPS, and there are few such applications in the very low energy (<50 eV) range, the fact that the TPP-2M formula cannot be used in the very low energy (<50 eV) range is still a shortcoming in some special applications. The TPP-LASSO-S formula [42] is a modified TPP-like formula derived by our team by machine learning (ML). Compared with the TPP-2M formula, the TPP-LASSO-S formula has increased accuracy with the introduction of Z and the power of ML but still has not overcome the inability of the empirical formula for IMFPs in the very low energy (<50 eV) range.

In contemporary materials science, ML is playing an increasingly important role because it can provide promising models for problems when a reliable empirical formula is not available [44]. Recently, many studies using ML on datasets or spectra [45,46] have shown the advantages of ML in materials science, thus guiding the methodology for application of ML. Therefore, in this paper, to overcome the unavailability of traditional empirical formulae on lower energies (<50 eV) and for ease of application, we consider using ML to provide a convenient way to calculate IMFPs. We established a robust database for IMFP and predicted unknown IMFPs of 18 transition and lanthanide metals. We show that the proposed ML scheme fits known IMFP data with an accuracy like or even better than that of the TPP-2M formula. Although the FPA-calculated IMFP values are not very accurate in the very low energy (<50 eV) regime due to the FPA calculational model, the trained ML model of IMFPs can be extended to the range even <10 eV, which can at least show illustrative trends. Moreover, we provide suitable descriptors for the training of the ML model over a broad energy range of the IMFP calculations, and the prediction ability of our algorithm is systematically discussed through leave-one-out cross-validation (LOOCV) and comparison with FPA-calculated IMFPs and experimental IMFPs that is not included in the training database. We also analyze the advantages and disadvantages of the used ML method and empirical formulae, since the ML method here shows good prediction power on materials with similar physical and chemical properties, while the empirical formulae, e.g., the TPP-2M formula, are effective on the prediction of unfamiliar material IMFPs. Finally, we discuss the possibility to directly extend the ML-predicted elemental IMFPs to compounds, thus revealing the future direction for ML methods used on IMFP prediction.

II. THEORETICAL METHOD AND RESULTS

The performance of ML models strongly depends on the database and training algorithm [47]. Through decades of study, researchers have accumulated numerous IMFP results that can serve as a reliable database to build a ML model. Shinotsuka *et al.* [24] theoretically computed IMFPs for 41 different elements with complete optical constant data over a wide energy range with the FPA. Here, these IMFP data were used as an initial database for the ML model. In this paper, the energy range and mesh are basically the same as in Ref. [24] because we used these data as a training and testing set. The detailed energy mesh is now listed in the Appendix, together with the prediction results of transition and lanthanide metals. It must be stated that Shinotsuka *et al.* did not publish their FPA-calculated IMFP values at very high (>200 keV) and very low (<50 eV) energies due to the limited accuracy in these energy regimes, as mentioned above, so data for these regions were communicated privately. That is, the information for 41 common elements were included in the ML model to quantitatively investigate the relationship between material parameters (Table I) and IMFPs [24]. For the consideration of both material and energy dependence of IMFP values, in the regression, there are 129 points dependent on electron energies for each material, so totally there are $129 \times 41 = 5289$ data instances in the dataset. We noted that all the electron energy values are above Fermi energy in the database, as well as the calculations, tables, and figures in this paper.

Another part of building the database for ML is to select proper input parameters, namely, the descriptors of the material parameters. Because the TPP-2M formula has achieved a good description of the IMFP, the parameters in the TPP-2M

TABLE I. Parameters used in ML, including atomic number (Z), atomic mass (M), density (ρ), number of valence electrons per atom (N_v), free-electron plasmon energy (E_p), bandgap energy (E_g), Fermi energy (E_F), and atomic radius (R).

Element	Z	M	$\rho(g/cm^3)$	N_v	E_p (eV)	E_g (eV)	E_F (eV)	R (pm)
Li	3	6.94	0.534	1	7.99	0	4.74	145
Be	4	9.01	1.848	2	18.44	0	14.3	105
C (graphite)	6	12.01	2.25	4	24.93	0	20.4	70
C (diamond)	6	12.01	3.515	4	31.16	5.5	20.4	70
C (glassy)	6	12.01	1.8	4	22.3	0	20.4	70
Na	11	22.99	0.971	1	5.92	0	3.24	180
Mg	12	24.31	1.738	2	10.89	0	7.1	150
Al	13	26.98	2.7	3	15.78	0	11.2	125
Si	14	28.09	2.33	4	16.59	1.1	12.5	110
K	19	39.10	0.862	1	4.28	0	2.12	220
Sc	21	44.96	2.989	3	12.86	0	5.8	160
Ti	22	47.87	4.51	4	17.68	0	6	140
V	23	50.94	6.11	5	22.3	0	6.4	135
Cr	24	52.00	7.14	6	26.14	0	7.8	140
Fe	26	55.85	7.874	8	30.59	0	8.9	140
Co	27	58.93	8.9	9	33.58	0	10	135
Ni	28	58.69	8.902	10	35.47	0	9.1	135
Cu	29	63.55	8.96	11	35.87	0	8.7	135
Ge	32	72.59	5.32	4	15.59	0.67	12.6	125
Y	39	88.91	4.469	3	11.18	0	4.4	180
Nb	41	92.91	8.57	5	19.56	0	5.3	145
Mo	42	95.94	10.28	6	23.09	0	6.5	145
Ru	44	101.07	12.41	8	28.54	0	6.9	130
Rh	45	102.91	12.41	9	30	0	6.9	135
Pd	46	106.42	12.02	10	30.61	0	6.2	140
Ag	47	107.87	10.5	11	29.8	0	7.2	160
In	49	114.82	7.31	3	12.59	0	4.82	155
Sn	50	118.71	7.31	4	14.29	0	5.51	145
Cs	55	132.91	1.88	1	3.43	0	1.73	260
Gd	64	157.25	8.23	9	19.77	0	3.5	180
Tb	65	158.93	8.25	9	19.69	0	4	175
Dy	66	162.50	8.78	9	20.08	0	3.5	175
Hf	72	178.49	13.31	4	15.73	0	7.9	155
Ta	73	180.95	16.65	5	19.53	0	8.4	145
W	74	183.85	19.3	6	22.86	0	10.1	135
Re	75	186.21	21.02	7	25.6	0	10.7	135
Os	76	190.23	22.61	8	28.08	0	11.4	130
Ir	77	192.22	22.65	9	29.66	0	11.2	135
Pt	78	195.08	21.45	10	30.2	0	10.6	135
Au	79	196.97	19.32	11	29.92	0	9	135
Bi	83	208.98	9.79	5	13.94	0	12.6	160

formula [24] were used, as listed in table 1 of Ref. [24]. Furthermore, we found that the inclusion of the atomic number Z and atomic radius R in the descriptors markedly improved the accuracy of the ML model, together with the other parameters listed in Table I. To describe the correlation between these descriptors, we used the Pearson's correlation coefficient r, which is defined as

$$r = \left| \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}} \right|, \tag{5}$$

where X_i and Y_i are two descriptors, and $\bar{X}$ and $\bar{Y}$ are the averaged values over n data points. According to Eq. (5), $r = 1$ indicates an exact linear correlation between X and Y, whereas $r = 0$ implies no correlation. Using the above parameter information and Eq. (5), the applicability of parameters to IMFP was analyzed as follows.

In this paper, the same descriptors, i.e., material-dependent physical parameters, were used as those employed in the TPP-2M equation. In Ref. [24], the TPP-2M equation included four material-dependent physical parameters to predict IMFPs: namely, atomic mass (M), density (ρ, in grams per cubic centimeter), number of valence electrons per atom (N_v), and bandgap energy (E_g, in electronvolts). Among these parameters, M and ρ are basic physical parameters, and E_g as well as the Fermi energy (E_F, in electronvolts), as basic material-dependent parameters, describe the basic model of

TABLE II. Detailed RMSD and its variance for different ML models appearing in Fig. 1(b).

ML algorithms	Testing set ratio	Average RMSD	RMSD variance
GLR	10%	37.98%	1.22×10^{-4}
SVM	10%	4.90%	2.66×10^{-6}
GPR	10%	0.69%	1.31×10^{-6}
GLR	30%	37.92%	4.10×10^{-5}
SVM	30%	4.92%	9.98×10^{-7}
GPR	30%	0.78%	8.24×10^{-7}
GLR	50%	37.87%	2.34×10^{-5}
SVM	50%	4.97%	5.35×10^{-7}
GPR	50%	0.93%	1.15×10^{-6}
TPP-2M		4.98%	

the energy band. The descriptors E_p and N_v are responsible for the density of valence electrons in a material. Here, N_v is usually associated with valence electrons but may include shallow core electrons in some materials. Also, E_p, which is another parameter that is related to the IMFP, is related to N_v by $E_p = 28.8(N_v\rho/M)^{1/2}$ (eV), which represents the oscillator strength for electrons that strongly contribute to the inelastic scattering. Figure 1(a) shows an obvious relationship between E_F and E_p. This relationship occurs because E_F and E_p are dependent on the electron number density for free-electron-like solids. However, this is not suggesting that E_F is not necessary in this paper, even if E_F is not appearing in the TPP-2M formula. In the theoretical model of calculating IMFP with optical constants, such as the FPA or Mermin method, E_F is a very important parameter. In the calculation of FPA, although E_F is not essential for predicting IMFPs at relatively high energies (>50 eV), it is very important in low energy IMFP calculations [48]. As was mentioned before, IMFP values are expressed as the energy above the Fermi energy in this paper, the same as the FPA-calculated database used here. In other words, electron energy (E') is from the bottom of the conduction band in the FPA calculations for a conductor, while the upper limit of the integration in FPA is $E' - E_F$, not E' [24], which is due to the Pauli exclusion principle. Therefore, the IMFP calculation with the FPA is associated with E_F. In the case of a high energy region of at least >50 eV, E_F can be ignored compared with the electron energy, so E_F does not apparently influence the calculation of the FPA. Inversely, the IMFPs in very low energies (<50 eV) are sensitive to the value of E_F. For the same reason, the TPP-2M formula does not contain E_F because its application energy region is >50 eV, while the ML model in this paper

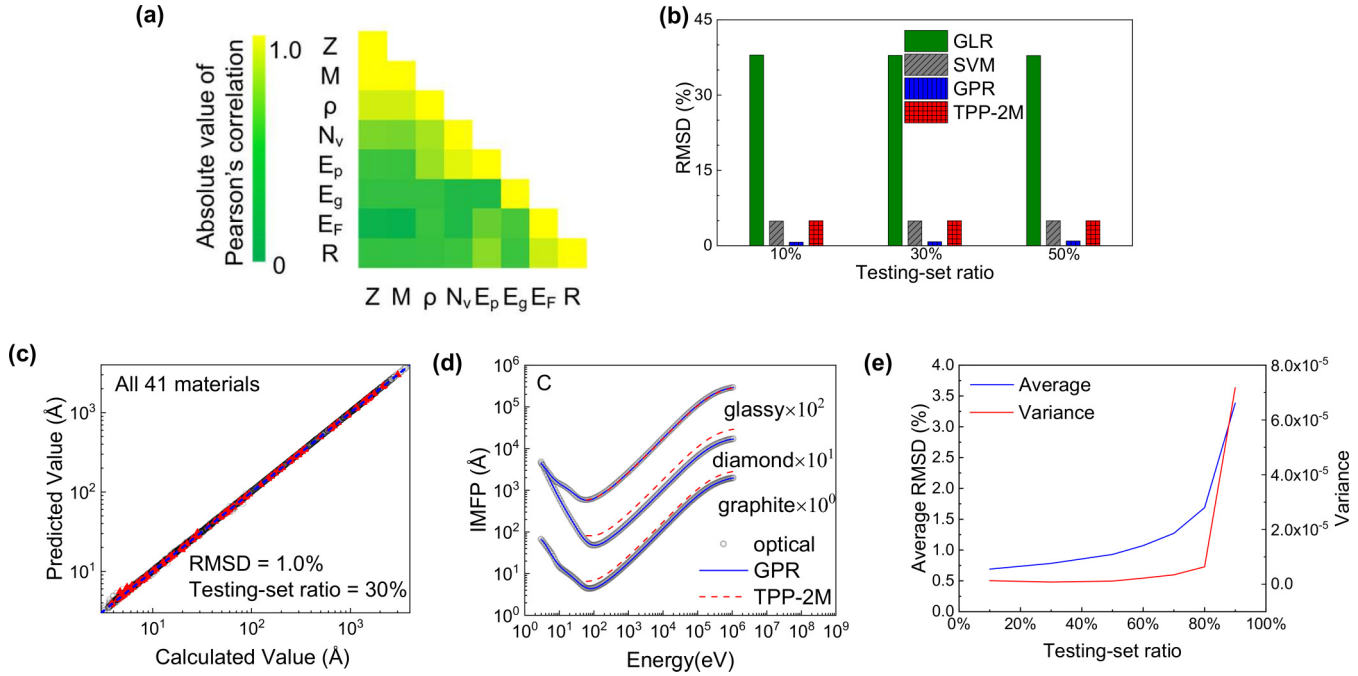

FIG. 1. (a) Correlations between adopted descriptors: atomic number (Z), atomic mass (M), density (ρ), number of valence electrons per atom (N_v), free-electron plasmon energy (E_p), bandgap energy (E_g), Fermi energy (E_F), and atomic radius (R). (b) The average root mean square deviation (RMSD) of different machine learning (ML) models. The models include Gaussian process regression (GPR), support vector regression (SVR), generalized linear regression (GLR), and the TPP-2M formula >50 eV for comparison. Detailed RMSD values and variances are shown in Table II. To ensure fair comparison, the calculation of RMSD with the TPP-2M formula also used Eq. (9). (c) Learning performance of inelastic mean free paths (IMFPs) for all materials. The x axis is the IMFP calculated using the full-Penn algorithm (FPA) in Ref. [24], and the y axis is the IMFP predicted using GPR. The energy range covers all IMFPs in the dataset. The blue line is a diagonal, which means agreement between predicted and calculated IMFPs, and the red triangles are predicted results. (d) Comparison between the GPR model, TPP-2M formula, and calculations with optical data. The blue solid line is IMFPs predicted by the GPR model, the red dashed line is IMFPs predicted by the TPP-2M formula, and the black dots are IMFPs calculated with optical data. The results for three typical carbon allotropes are shown. The electron energies are expressed with respect to the Fermi level. (e) The average RMSD and variance of the trained GPR model. The model for each testing set ratio is trained 100 times separately by changing the random partition of training and testing sets.

estimates IMFPs at very low energies (<50 eV). Therefore, in the ML model, we must adopt E_F as a basic parameter. Figure 1(a) also reveals that the correlations among Z, M, and ρ are the highest: these descriptors are the most basic physical parameters when predicting IMFPs, which means that they cannot be considered as repeated features and must be retained. In addition, N_v changes periodically with increasing atomic number and has a weak correlation with the basic physical parameters R (in picometers), Z, M, and ρ. Here, E_p is correlated with N_v, ρ, and M. As for E_F and E_g, because they are parameters in band theory, they have no direct correlation with the other descriptors. These physical quantities have already been included in the TPP-2M formula or FPA and thus must also be included here. As shown in Fig. 1(a), the descriptors used in this paper show weak correlations with each other. Therefore, because they are almost independent of each other, all are necessary for the ML model.

There are various training algorithms for the ML model with the ability to recognize patterns from a dataset and then use these patterns to make predictions for new data [49]. Here, we compare the performance of three regression algorithms for the ML model of IMFPs: generalized linear regression (GLR), Gaussian process regression (GPR) [50], and support vector regression (SVR) [51].

(i) GLR

Linear regression is often performed using the least-squares method, producing a linear relationship between descriptors and a target. GLR is not like simple linear regression, which gives a response with a certain distribution, e.g., a Poisson distribution. The distribution is a function added between a simple linear regression and the final regression results.

(ii) GPR

The GPR model is a probabilistic model belonging to a generic supervised learning method. The GPR model provides a probabilistic distribution of a new output value by the descriptors $P(y_{\text{new}}|y_{\text{train}}, x_{\text{train}}, x_{\text{new}})$ based on the training result for each step. In the step-by-step optimization, the joint distributions of the regressed function follow a Gaussian process:

$$f(x) \propto \text{GP}[0, k(x, x')], \tag{6}$$

where $k(x, x')$ is the kernel function. This kernel function is related to the shape of the target, which is a radial basis function in this paper:

$$k(x, x') = \exp\left[-\frac{1}{2}d\left(\frac{x}{l}, \frac{x'}{l}\right)^2\right], \tag{7}$$

where l is the length scale parameter, and $d(a, \text{b})^2$ is the mean square derivative of a and b. For available distributions of the function and targets, the posterior distribution of the adjusted function is calculated through a Gaussian process.

(iii) SVR

The essential thought of SVR is to transform a regression into a linear regression. SVR involves reflecting the descriptors into a high-dimensional feature space in which a high-dimensional linear regression can be performed. The kind of SVR used in this paper is ε-SVR, which allows a tolerance gap ε between the true target and the learning target. Within the gap, the result will be an acceptable result. The actual optimization uses quadratic programing algorithms. The decision function is

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)k(x, x') + b, \tag{8}$$

where α_i, $\alpha_i{}^*$ are Lagrange multipliers, and $k(x, x')$ is a kernel function, as shown in Eq. (7).

To prepare the training data, the database was divided into a training dataset and testing dataset. The ML model was trained with the training set, and then the accuracy of the model was calculated with the testing set. The testing set ratio was used to indicate the size ratio of the testing dataset over the total database. A larger testing set ratio indicates the model was trained with a smaller training dataset. Figure 1(b) shows the performance of three ML algorithms for different datasets with three testing set ratios and the detailed RMSD values and variances are shown in Table II. Each ML algorithm was trained 100 times by changing the random partition of training and testing sets in the sight of credibility. The performance of the TPP-2M formula for applicable energies (>50 eV) is included as a reference. The accuracy of the model was measured by the root mean square deviation (RMSD) as

$$\text{RMSD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\log\lambda_{\text{pred}}(E_i) - \log\lambda(E_i)}{\log\lambda(E_i)}\right]^2}, \tag{9}$$

where n is the total number of data points in the testing set, E_i is the electron energy, $\lambda_{\text{pred}}(E_i)$ is the predicted IMFP, and $\lambda(E_i)$ is the target value calculated by the FPA. The closer that the RMSD is to zero, the better the prediction. Figure 1(b) reveals that the GLR has the worst performance in the different testing set ratios. The results of the GPR and SVR are much better and are close to the performance of the TPP-2M formula with the high-energy data. GPR showed the best performance of the three algorithms. Therefore, we used this algorithm to train the ML model. For the testing set ratio, we chose it as 30%, which is widely used in the ML community [52]. This is also allowed with Fig. 1(e). In this figure, we expand the ratio of the testing set until 90% for GPR and trained every model 100 times. For a testing set ratio >30%, the average RMSD and its variance is rising. Above all, according to this evidence, our chosen testing set ratio was 30%.

Figure 1(c) presents the training results of IMFP with the GPR method determined above. There is good agreement between the GPR-predicted value and the standard values calculated by the FPA. Figure 1(d) compares the IMFPs predicted with the GPR method with those obtained from the TPP-2M formula using optical data for three carbon allotropes. Our ML approach shows remarkable performance even if the predictions of carbons are typically poor among all 41 materials, whereas the TPP-2M equation cannot achieve a balanced description accuracy for three carbon allotropes, despite the isolated good agreement for glassy carbon. These results reveal that our ML method is much stronger than the TPP-2M equation in terms of IMFP data retrieval. Additionally, during this procedure, the critical features mentioned above were selected. In fact, during the regression of the GPR

TABLE III. Length scales of parameters (electron energies excluded) used in ML optimized by GPR.

Element	Z	M	$\rho(g/\mathrm{cm}^3)$	N_v	E_p (eV)	E_g (eV)	E_F (eV)	R (pm)
Maximum value	83	208.98	22.65	11	35.87	5.5	20.4	260
Length scale	15970	103	1220	8.95	11.4	6.24	11.8	76.5

algorithm, the optimized value length scale l for each feature x appearing in Eq. (7) was also given by the program. According to Eq. (7), a larger length scale means a smaller deviation of kernel function, thus leading to a weaker impact to the target value. In other words, the value of length scale can reflect the relativity of each feature toward the target value. Through the different testing set ratio attempts, the features were selected by checking the value of the optimized length scale. In the procedure of GPR, many other features including Z and R, for example, and electrical and thermal conductivity were tried, but their optimized length scale usually reached the upper bound of the optimization, in other words, show low impact to IMFPs. Optimized length scales of the features used in ML are shown in Table III. In the features used, the length scales are all comparable with the maximum value of each feature, while the length scale of Z and ρ is relatively larger. In fact, this is caused by the introduction of Z and R. For Z, it has an obvious linear relationship with M, as shown in Fig. 1(a); for R, it is related to atomic effective volume ($\sim R^3$), which is also linear with M/ρ. On one hand, M and ρ appeared in the TPP-2M formula, which is already being used to describe IMFPs; on the other hand, the necessity of Z and R was proven with a data-driven idea, by regressing another powerful empirical formula with LASSO [42] in a previous paper. Although these features are not independent to each other, which means the length scale of these parameters will inevitably include large values, they are very easy to obtain because they are all basic features (material-dependent parameters) in the periodic table. It is harmless to include them in the selection of features. Specifically, the TPP-2M formula does not include Z and R yet; we stated that they are necessary in GPR. Meanwhile, many other features were deleted from the candidates because their length scales were not reasonable, and finally, the eight features used in this paper were decided.

We now demonstrate the ability of the current ML model to predict IMFPs. In the ML area, one of the algorithms to monitor and avoid overfitting is cross-validation (CV). In this paper, we also introduced one cross-validation, e.g., LOOCV, which is also widely used to test the prediction performance of ML algorithms. In the LOOCV method, we trained the ML

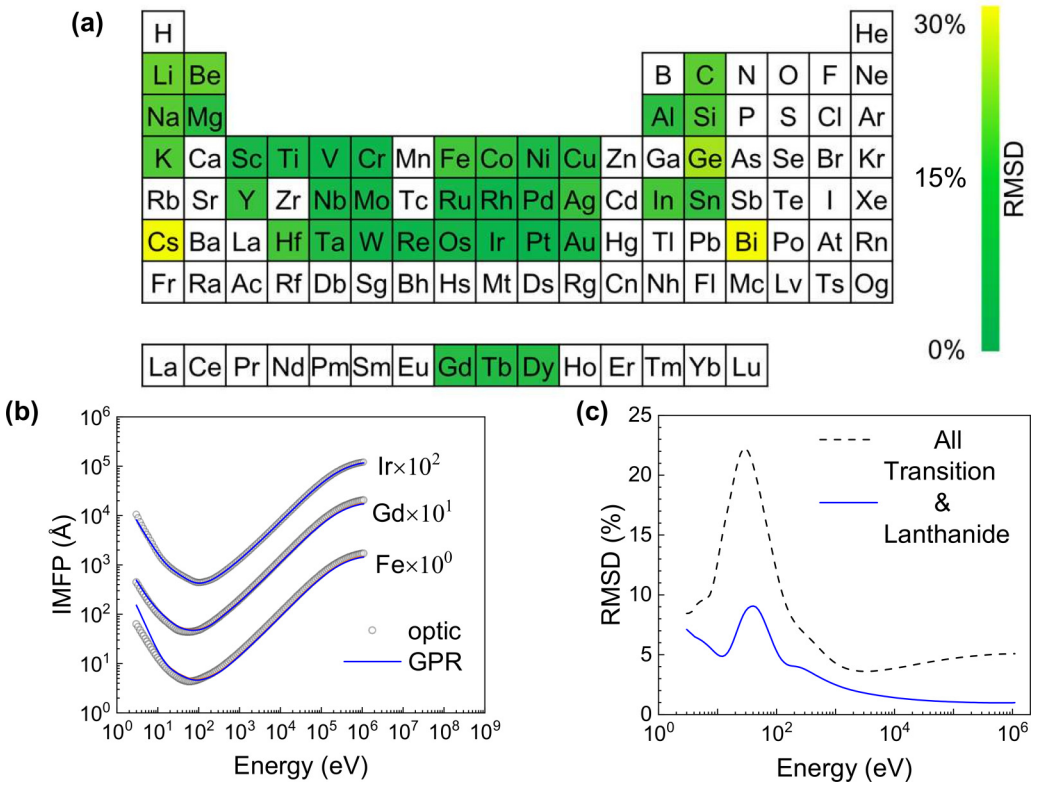

FIG. 2. (a) Direct representation of root mean square deviations (RMSDs) determined for the 41 investigated materials in the periodic table. The gradual color change from green to yellow corresponds to the RMSD value. A detailed RMSD value distribution is provided in Table IV. (b) Some typical leave-one-out cross-validation (LOOCV) results (Ir, Gd, and Fe); the electron energies are expressed with respect to the Fermi level. The variances are also shown in the figure, which are very small. (c) RMSD as a function of energy for different kinds of materials. The black dashed line is the average RMSD for all materials. The blue solid line is for transition and lanthanide metals.

TABLE IV. RMSDs (%) of LOOCV results.

Element	Li	Be	C (graphite)	C (diamond)	C (glassy)	Na	Mg	Al	Si	K	Sc	Ti	V	Cr	Fe	Co	Ni	Cu	Ge
RMSD	12.37	12.29	8.10	17.02	7.98	10.59	5.12	5.29	9.59	9.74	2.93	3.11	1.74	1.67	7.88	6.10	3.17	4.31	19.68
Element	Y	Nb	Mo	Ru	Rh	Pd	Ag	In	Sn	Cs	Gd	Tb	Dy	Hf	Ta	W	Re	Os	Ir	Pt	Au	Bi
RMSD	6.62	2.66	2.28	2.52	1.70	2.96	6.69	8.49	6.00	28.59	4.49	4.10	4.46	8.09	3.97	1.87	1.71	2.27	1.36	1.10	2.57	28.89

model with 40 of the 41 materials. The performance of the model was then tested on the 41st material. In other words, the information of the 41st material was the testing set, and the data for the other 40 materials were the training set for the model. We repeated the process 41 times so that the prediction performance (characterized by RMSD) was obtained for all the materials in the current dataset. The LOOCV method was conducted as follows in this paper. First, all the data for a single material were taken out as a testing set, and then the data for the other 40 materials were used as a training set. Second, an ML process was run on the training set, and then the learning result was tested using the testing set. This process was repeated 41 times (once for each material) to generate a total of 41 cross-validation learning results. Because each training cycle is equivalent to using the data of 40 materials to predict the remaining one material, this method can effectively test the ability of GPR to predict the IMFP of a material.

Figure 2(a) displays a color map of the RMSDs of the 41 materials determined from the LOOCV testing (Table IV). The total average RMSD was 6.9% and >80% of the elemental solids had RMSDs within 10%. Careful examination of Fig. 2(a) reveals that the predicted IMFPs of the transition and lanthanide metals show very good accuracy with an average RMSD of only 3.5% for all the transition and lanthanide metals in the dataset. Figure 2(b) shows three typical LOOCV results with different accuracies for the transition or lanthanide metals Ir (1.36%), Gd (4.49%), and Fe (7.88%). Even the LOOCV results for Fe in Fig. 2(b) show satisfactory agreement with the results from optical data. The energy dependence of the RMSD values is presented in Fig. 2(c). For the transition and lanthanide metals, the ML model shows slightly larger RMSDs in the lower energy range, especially in the region from approximately 10 to 100 eV, and then gradually decreased as the energy increases. Considering the information in Figs. 2(b) and 2(c), we can draw the conclusion that the predictive performance of the GPR for transition and lanthanide metals is consistent and superior to that for other materials. Figure 2(c) also shows the typical RMSD distribution of transition and lanthanide metals with energy, which confirms the good prediction accuracy of the ML model for transition and lanthanide metals.

Figure 2(a) reveals that the total RMSDs for three elements exceeded 20%: namely, Ge (20%), Cs (29%), and Bi (29%). These large prediction biases probably originated from the lack of training data, especially data for elements with similar physical properties. For example, Cs and Bi are quite isolated from the other elements in the training set on the periodic table, as shown in Fig. 2(a). In addition, Ge has no neighboring elements in the periodic table that are in the training dataset. Another possible reason for the large prediction biases is that the current descriptors in the ML model may not be sufficient to describe Ge, Cs, and Bi well. Further investigations should be conducted to collect more training data and search for more universal descriptors for the IMFP data.

In Fig. 2(c), the energy dependence of total RMSD for all materials shows a similar trend to that of the transition and lanthanide metals, except that the absolute values were larger, especially in the very low energy (<50 eV) range. The highest RMSD was 22% at an energy of ∼30 eV. While this prediction error is nonnegligible, in addition to the inaccuracy of FPA in the low energy region (<50 eV), namely, our training set stated earlier, three points should be emphasized: (1) The large deviation in the very low energy (<50 eV) range mainly arose from the small number of elements in the dataset. For example, the RMSDs for the alkali and alkaline earth metals are much larger than those of the transition and lanthanide metals. This result occurs mainly because the difference of the physical properties between alkali and alkaline earth metals is relatively large, whereas other materials, e.g., the transition and lanthanide metals, are fairly similar to each other. (2) Barely any empirical formula has prediction power for low energy IMFPs. On one hand, the TPP-2M formula was not intended for use in the low energy region (<50 eV); on the other hand, despite the empirical formula including the very low energy (<50 eV) range has been achieved in Ref. [43], there are still significant limitations for its application, as mentioned before. When the energy is as low as 50 eV, there are few channels for electrons to lose energy; meanwhile, the probabilities for inelastic scattering generally decrease for electron energies less than several times the excitation energy, which simultaneously causes the IMFP to increase in the very low energy (<50 eV) region. Therefore, the calculation of FPA in the very low energy (<50 eV) is not as accurate as intermediate energies (>50 eV or <200 keV) because of the neglect of many effects. The imprecise values of IMFPs in the very low energy (<50 eV) as a training set may mislead the prediction of ML, thus resulting in a large peak in low energy region. Despite the training set, namely, FPA-calculated IMFPs from Ref. [24], has less validity at extreme energies (<50 eV or >200 keV) than the intermediate energy region and may cause lower reliability in our GPR prediction, it is not a disadvantage for GPR. (3) From the IMFP calculation methods (e.g., FPA), we can see that the description of IMFP in the low and high energy regions relies on different material-dependent parameters (features), respectively. However, to meet the need for all the energy regions in this paper, the features were selected together. This may lead to a poor prediction for the low energy region. In other words, as was mentioned before, the IMFP calculation using FPA is directly associated with ELF. The shape of ELF is quite different in lower energies (<100 eV) for each material.

TABLE V. Properties of transition and lanthanide metals used to predict IMFPs. Note that these are materials whose IMFPs cannot be calculated by physical theory (e.g., FPA) because of a lack of ELF or optical constants.

Element	Z	M	$\rho(g/\mathrm{cm}^3)$	N_v	E_p (eV)	E_g (eV)	E_F (eV)	R (pm)
Mn	25	54.94	7.47	7	28.10	0	10.9	140
Zn	30	65.38	7.14	12	32.97	0	9.47	135
Zr	40	91.22	6.51	4	15.39	0	5.8	155
Tc	43	98.00	11.50	7	26.10	0	7.1	135
Cd	48	112.41	8.65	12	27.68	0	7.47	155
La	57	138.91	6.15	9	18.17	0	3.7	195
Ce	58	140.12	6.69	9	18.88	0	2.9	185
Pr	59	140.91	6.64	9	18.76	0	3.8	185
Nd	60	144.24	7.01	9	19.05	0	4	185
Pm	61	145.00	7.26	9	19.34	0	4.2	185
Sm	62	150.36	7.35	9	19.11	0	4.5	185
Eu	63	151.96	5.24	9	16.05	0	4.2	185
Ho	67	164.93	8.80	9	19.95	0	4.9	175
Er	68	167.26	9.07	9	20.12	0	3.9	175
Tm	69	168.93	9.32	9	20.30	0	4.1	175
Yb	70	173.05	6.57	9	16.83	0	2.7	175
Lu	71	174.97	9.84	9	20.49	0	6	175
Hg	80	200.59	13.53	12	25.91	0	7.13	150

We noted that the used features, even for E_p and E_g, cannot completely show the characteristics of ELF in these energies. Therefore, in sight of the energy-dependent degree, the GPR prediction ability in this energy region, especially from material-dependent parameters, is inevitably poorer. Based on the specified database, the current ML model covers both the low and high energy regions, which means that it is a reliable approach and can be applied to the prediction of the IMFP.

With our confidence in the current ML model established, especially for transition and lanthanide metals, we then made predictions of unknown IMFPs using the parameters listed in Table V. Figure 3 shows the predicted IMFP curves for Mn, Zn, Zr, Tc, Cd, and Hg as representative examples. For comparison, we also include the IMFPs predicted by the TPP-2M formula, which only works in the high energy region. Because of the robustness of our ML method, the trend predicted by the ML approach is consistent for the six materials, and the high energy region agrees well with the results from the TPP-2M formula. The most unagreed result in the high energy region between our ML approach and the TPP-2M formula are the materials Zn and Hg. We noticed that Zn and Hg both have just one neighbor material on the degree of period table in the training data [see Fig. 2(a)], but other materials in Fig. 3 have both neighbors. Additionally, Hg is the only liquid metal existing in simple materials, and this fact may bring out some unique trends on the IMFP curve. These trends may not have been seized by FPA calculation as our training data; thus, ML should have a larger deviated result. Considering that our training set includes the FPA-calculated IMFP data, and the TPP-2M formula is also derived from FPA-calculated IMFP data, the agreement between our ML data and TPP-2M formula is very reasonable. These results demonstrate that the ML method is very reliable because it is only dependent on the reliability of the input data; it does not require any artificial or subjective factors that are often included in empirical formulas. More details of the ML prediction data used to determine the IMFPs of transition metals, including lanthanide metals [uncolored elements in Fig. 2(a)] are described in the Appendix.

However, the only comparison with the empirical formula is not pursuable. Here, another series of FPA calculations for materials not included in Ref. [24], i.e., the training set of our ML approach, is carried out for further comparison. The most sensitive factor in the FPA calculation is ELF. In fact, the reason why Shinotsuka *et.al.* did not include several materials in the FPA calculation [24] is that they could not find suitable ELFs at that time for Mn and Zr, etc. Suitable means their quality, namely, two sum rules have large errors.

Two sum rules, i.e., the oscillator strength sum rule (f-sum rule) and the perfect screening sum rule (ps-sum rule) [53], were applied to check the accuracy of the ELFs of Mn and Zr used in this paper. The f-sum rule Z_{eff} is given by

$$Z_{\mathrm{eff}} = \frac{2}{\pi \Omega_P^2} \int_0^{\omega_{\max}} \omega \mathrm{Im}\left\{\frac{-1}{\varepsilon(\omega)}\right\} d\omega, \tag{10}$$

where $\hbar\Omega_P = \sqrt{4\pi n_a e^2/m}$, $n_a = N_a \rho / m_e$ is the number density of atoms, N_a is the Avogadro's number, ρ is the mass density, and m_e is the atomic weight. The ps-sum rule P_{eff} can be obtained from the Kramers-Kronig relation as Ref. [53]:

$$P_{\mathrm{eff}} = \frac{2}{\pi} \int_0^{\omega_{\max}} \frac{1}{\omega} \mathrm{Im}\left\{\frac{-1}{\varepsilon(\omega)}\right\} d\omega + \mathrm{Re}\left\{\frac{1}{\varepsilon(0)}\right\}, \tag{11}$$

where $\mathrm{Re}\{1/\varepsilon(0)\} = 0$ for conductors. The theoretical values of Z_{eff} and P_{eff} are atomic number and unity, respectively, in the limit of $\omega_{\max} \to \infty$. In this paper, $\omega_{\max}$ was 1 MeV.

After a lot of literature research, we validated the amount of ELFs of many simple materials and finally found two suitable ELFs for two materials, Mn and Zr. The ELFs of Mn used in the calculation of the IMFP were taken from Adachi [54]

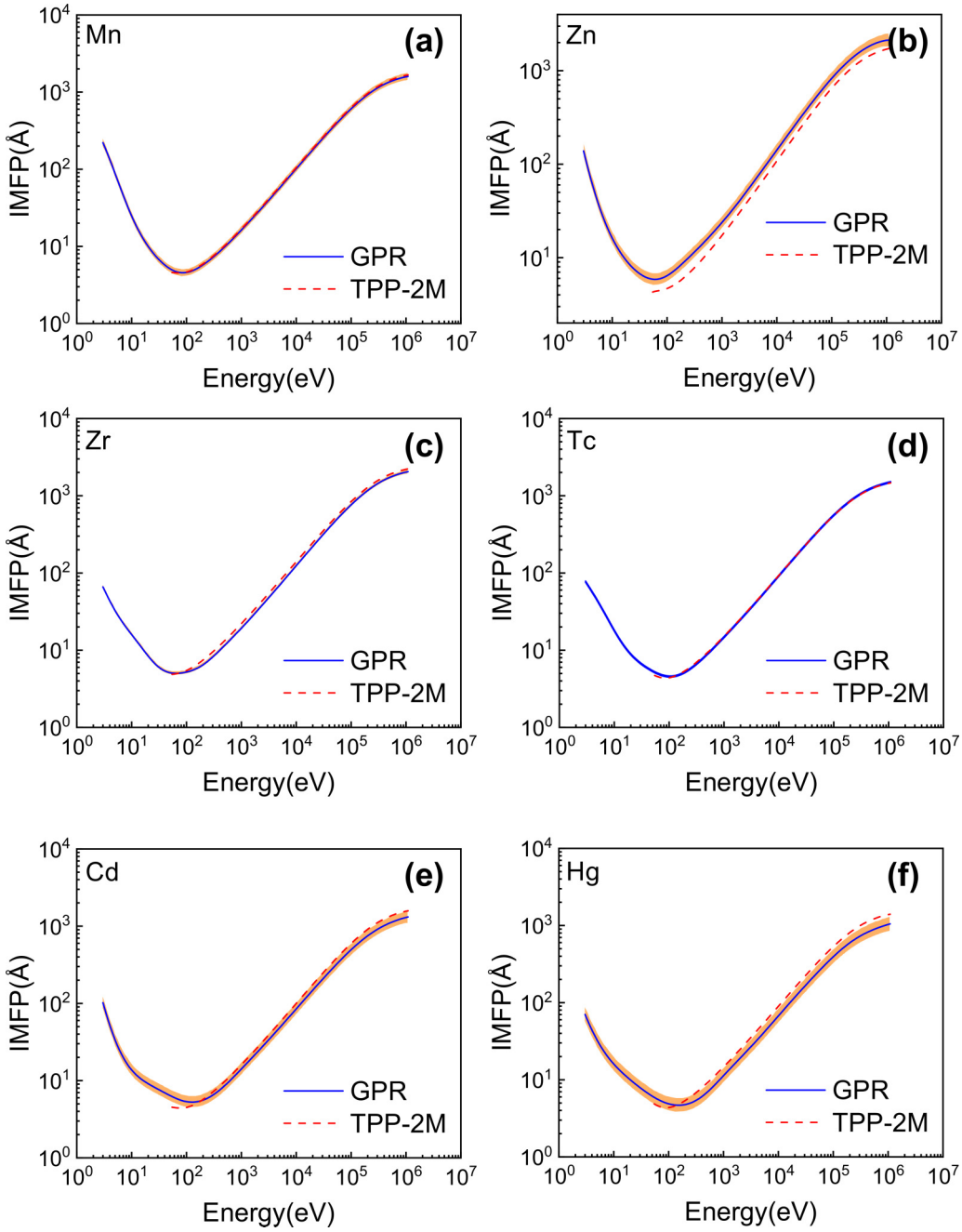

FIG. 3. Representative predictions for the transition metals Mn, Zn, Zr, Tc, Cd, and Hg. Blue solid curves are inelastic mean free paths (IMFPs) predicted by Gaussian process regression (GPR), red dashed curves are IMFPs predicted by the TPP-2M formula. All electron energies are expressed with respect to the Fermi level. The variances are also shown in the figure.

in the photon energy range of 0.07–6.6 eV, from Wehenkel and Gauthé [55] in the range of 7.0–110 eV, and from Henke *et al.* [56] in the range of 0.11–30 keV [see Fig. 4(a)]. The ELFs of Zr were taken from Prieto *et al.* [57] in the photon energy range of 0–80 eV and from Henke *et al.* [56] in the range of 0.08–30 keV [see Fig. 4(b)].The ELFs for energy losses between 0.03 and 1 MeV were calculated from atomic scattering factors [58]. The discontinuity of ELFs shown in Figs. 4(a) and 4(b) is because the ELFs are composed of multiple sets of experimental data, and the results of different experiments are different, including the connective energy points. Figures 4(c)–4(f) show the f- and ps-sum rule checks of the ELFs of Mn and Zr used in this paper. Table VI lists the results of the f- and ps-sum rules of the ELFs of Mn and Zr. In the validation of sum rules for both materials, the ps-sum rule shows very small relative errors; the f-sum rule shows

TABLE VI. List of f- and ps-sum rule checks of ELFs of Mn and Zr.

	f-sum rule	Relative error	ps-sum rule	Relative error
Mn	21.47	−14.1%	0.984	−1.6%
Zr	37.24	−6.9%	1.016	1.6%

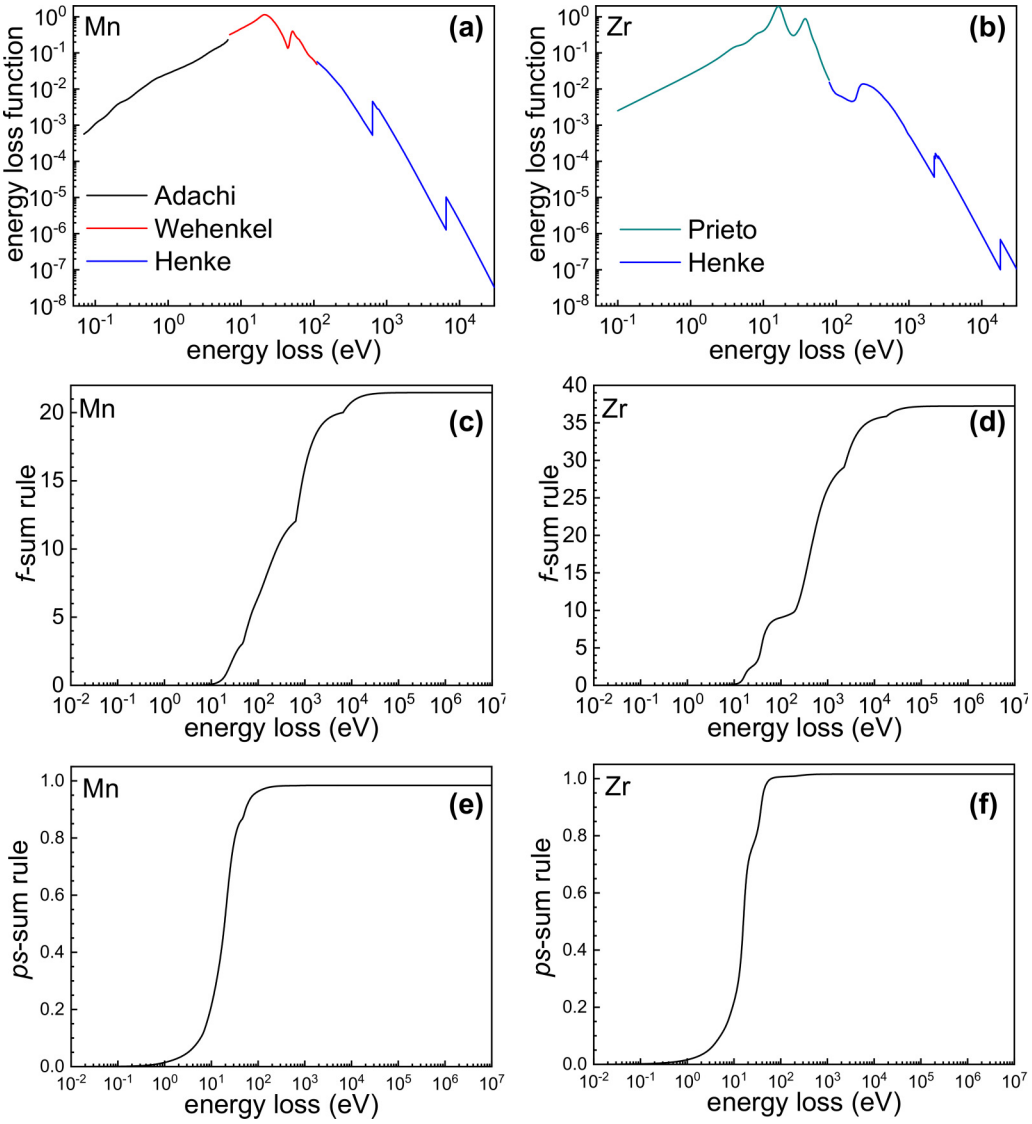

FIG. 4. Energy loss functions (ELFs) and two sum rules for Mn and Zr.

relatively larger error, especially for Mn, but still acceptable comparing with other ELFs we found.

Through the validated ELFs of Mn and Zr, IMFPs were calculated through FPA. Figure 5 shows the comparison of FPA-calculated with GPR-predicted IMFPs on Mn and Zr. In the low energy region (<100 eV), the GPR-predicted IMFPs for Mn are larger than the FPA result but still illustrate the trend; in the high energy region (>100 eV), the GPR-predicted IMFPs are like the FPA results but slightly lower. Meanwhile, for Zr, the relative position of the curves for GPR and FPA are almost similar with the situation of Mn in the whole energy region, but Zr has a better GPR result than Mn. The larger deviation of Mn is highly probable due to the underestimated ELF used in the FPA calculation. As shown in Fig. 4(c) and Table VI, the f-sum rule of the ELF for Mn has a relative error of -14.1%, representing the ELF of Mn used herein has apparent error with the true value, especially for the high energy part. This result shows that, in the future, we are willing to use the GPR-predicted IMFP as a reference for splicing different experimental ELFs between different energy segments.

Moreover, to compare the predictive power of empirical formulae with GPR, we also plot the IMFPs predicted by various empirical formulae including the TPP-2M formula in Fig. 5 for the whole energy region. The TPP-LASSO-S formula was also developed by our group in Ref. [42], while the S1 and G1 formulae were developed by Seah and Gries [38]; details can be found in Ref. [59]. Although the applicable energies for the TPP-2M formula are >50 eV and for TPP-LASSO-S, G1, and S1 formulae >200 eV, the IMFPs predicted by each formula are shown in the whole energy region. The shorted values in the lower energy for each formula are because they give out negative IMFPs which cannot be shown in the figure. Firstly, the formulae and GPR show consistency with FPA-calculated data in high energies (>50 eV). To quantify the error, the relative deviation between GPR-predicted and FPA-calculated IMFPs, as well as the most representative one among all formulae, and the TPP-2M formula-predicted and FPA-calculated IMFPs are calculated in the high energy region (>50 eV). For Mn, the RMSDs [see Eq. (9)] of GPR and the TPP-2M formula are 3.77% and

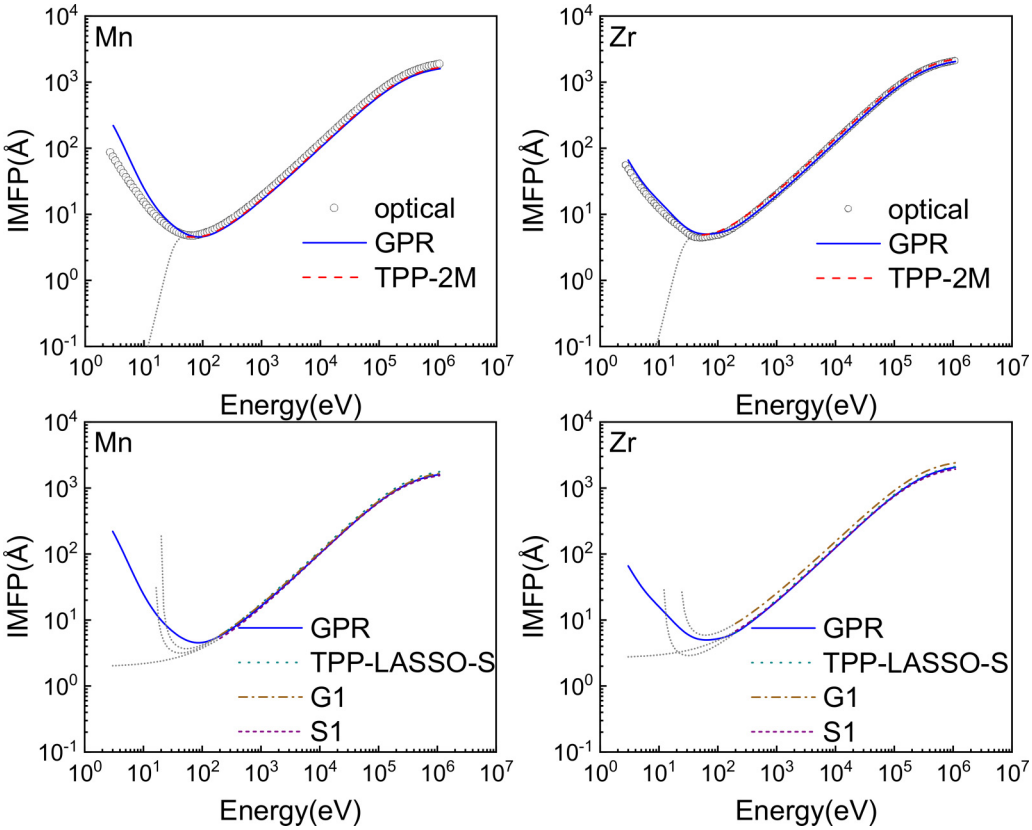

FIG. 5. Representative predictions for Mn and Zr. Blue solid curves are inelastic mean free paths (IMFPs) predicted by Gaussian process regression (GPR), red dashed curves are IMFPs predicted by the TPP-2M formula, the green dotted curves are IMFPs predicted by the TPP-LASSO-S formula, the brown dash-dotted curves are IMFPs predicted by the G1 formula, the purple short dashed curves are IMFPs predicted by the S1 formula, and black hollowed dots are IMFPs obtained from full-Penn algorithm (FPA) calculation. Despite the applicable energies (>50 eV for the TPP-2M formula, >200 eV for the other formulae), the predicted IMFPs for these energies for each formula are still shown in gray dots for comparison. For clarity, the comparison of each material is shown in two plots, respectively. All the electron energies are expressed with respect to the Fermi level.

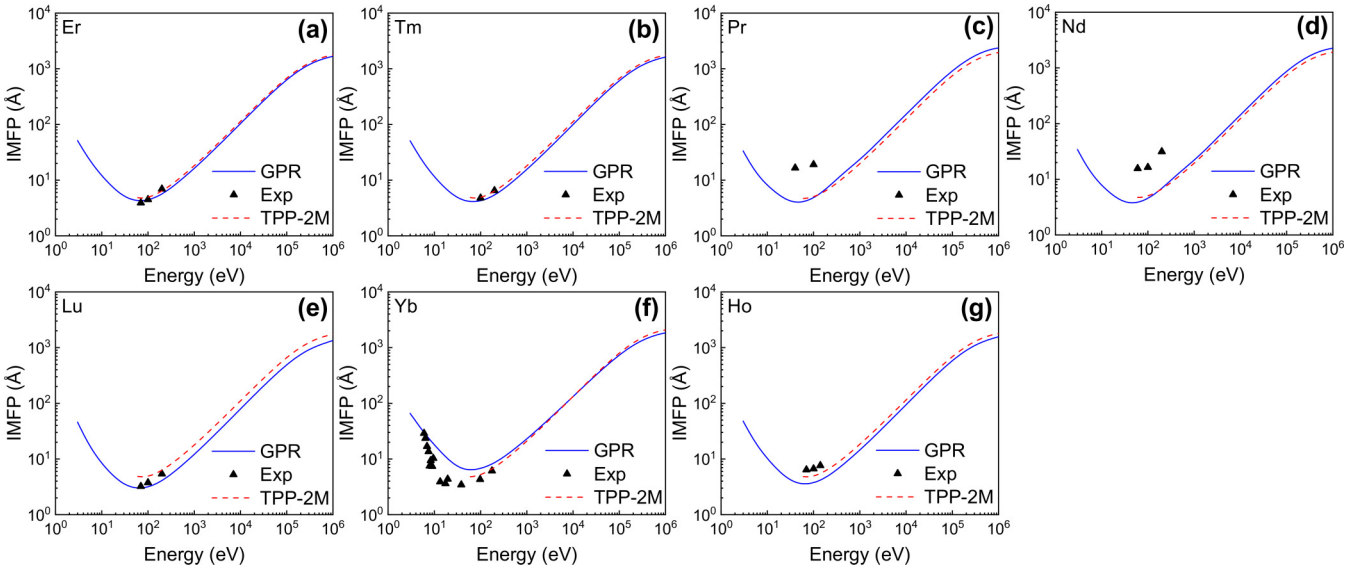

FIG. 6. Representative predictions for the transition and lanthanide metals Er, Tm, Pr, Nd, Lu, Yb, and Ho. Blue solid curves are inelastic mean free paths (IMFPs) predicted by Gaussian process regression (GPR), red dashed curves are IMFPs predicted by the TPP-2M formula, and black dots are IMFPs calculated from experimental data for comparison.

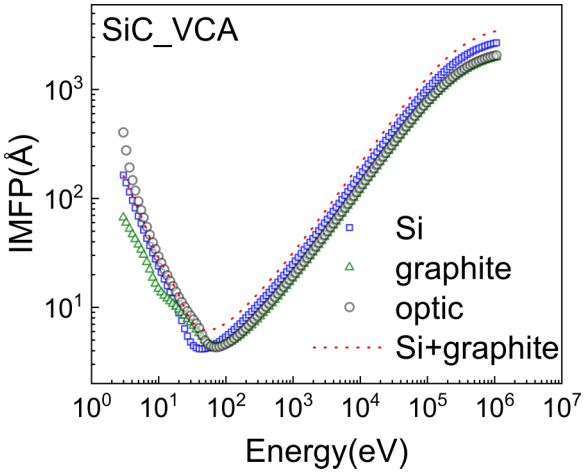

FIG. 7. Predictions for SiC based on elemental materials using the virtual crystal approximation. Blue square dots are inelastic mean free paths (IMFPs) of Si, and green triangle dots are IMFPs of C (graphite) calculated from experimental data in Ref. [24]; they are used for SiC IMFP calculation in this figure. Red dotted curve is IMFPs predicted by Si and C (graphite). Black dots are IMFPs calculated from experimental data in Ref. [25] for comparison.

2.91%, respectively, while those of Zr are 2.28% and 3.89%. Meanwhile, for the situation of energy >200 eV, the RMSDs for GPR, TPP-LASSO-S, G1, and S1 are 5.68%, 11.91%, 18.72%, and 26.18% for Mn, while those of Zr are 5.91%, 12.43%, 19.58%, and 27.46%, respectively. These relative deviations clearly indicate that, at least in the situation of Mn and Zr, GPR has the same predictive power compared with TPP-2M for electron energies >50 eV, and generally better than the TPP-LASSO-S, G1, and S1 formulae for electron energies >200 eV, about which researchers are frequently concerned. Additionally, in lower energies (<50 eV), even the trends of empirical formulae do not give certain reference values, while our GPR-predicted IMFPs are the most closed curve to FPA-calculated values. In fact, although there is no further experimental evidence for IMFPs of Zn, Tc, and Cd, we believe that only for the cases discussed in this paper, namely, from the known IMFP database to predict the IMFPs unknown for sporadic transition metals in the database, GPR can give out reliable results in the degree of statistics.

GPR is different from traditional empirical formulae for the different focus points. GPR-defaulted material features follow the Gaussian distribution, which means that the predictions of GPR focus on local information between similar materials in the prediction; but empirical formulae, including the TPP-2M formula, must contain the information of all the materials, including 14 organic compounds [23], 41 element materials [24], and 42 inorganic compounds [25]; thus, the IMFP descriptions for general materials must make a sacrifice for materials with special properties to contain all the material IMFPs. This characteristic can also be seen from Fig. 2(a): transition metals and lanthanides show a better prediction (lower RMSD), while the RMSDs for Bi and Cs are very large. Not only associated with atomic numbers, for the elementary materials far away in the periodic table, various physical and chemical properties may have large differences. Therefore, for GPR, the prediction of these materials with large differences in features will lack effective information, leading to poor learning effects of these materials, namely, large RMSDs. Therefore, comparing with GPR, the TPP-2M empirical formula obtained by mass data analysis is more suitable for prediction of the material IMFPs with less correlation with the known IMFP materials. Naturally, the "correlation" here means the feature values difference between materials (i.e., the correlation between physical and chemical properties). The ML method, e.g., GPR in this paper, has a stronger predictive ability when the known IMFP materials are more relevant to the material whose IMFPs need to be predicted.

Because TPP-2M is an empirical formula, it is appropriate to prove the accuracy of the prediction data with experiments in addition to the TPP-2M formula in Fig. 3. Figure 6 compares the ML results and those obtained from other calculations using the experimental data from Ref. [60], software ELSEPA developed in Ref. [61], and the calculation method inspired by Refs. [4,62–64]. The EAL λ_{EAL} [4] is another important physical quantity in surface analysis [64]. From the EAL data, an experimental IMFP value can be evaluated. EAL can be calculated as follows [62]:

$$\lambda_{\mathrm{EAL}} = \frac{x}{\ln(I^{\mathrm{S}}/I^{\mathrm{B}} + 1)}, \tag{12}$$

for all the predicted materials. In Eq. (12), x is the surface layer thickness and I^S/I^B is determined from the experimental data in Ref. [60]. EAL was then converted to IMFP to compare results. Using the software ELSEPA [61], the transport cross-section σ_{tr} was obtained. A linear interpolation was used to fit the energies. According to Ref. [63],

$$\lambda_{\mathrm{tr}} = (M'\sigma_{\mathrm{tr}})^{-1}, \quad M' = N_0\rho/A_m, \tag{13}$$

where M' is the atomic density, N_0 is the Avogadro number, A_m is atomic weight, and ρ is density. Equation (13) allows σ_{tr} to be converted to transport mean free path (TMFP) λ_{tr}. As described in Ref. [64],

$$\lambda_{\mathrm{EAL}} = (1 - 0.738\omega)\lambda_{\mathrm{IMFP}}, \omega = \frac{\lambda_{\mathrm{IMFP}}}{\lambda_{\mathrm{IMFP}} + \lambda_{\mathrm{tr}}}. \tag{14}$$

Using the known TMFP, IMFP can be determined according to

$$\lambda_{\mathrm{IMFP}} = \frac{(\lambda_{\mathrm{EAL}} - \lambda_{\mathrm{tr}}) + \sqrt{(\lambda_{\mathrm{EAL}} - \lambda_{\mathrm{tr}})^2 + 1.048\lambda_{\mathrm{EAL}}\lambda_{\mathrm{tr}}}}{0.524}. \tag{15}$$

Therefore, Eq. (15) allows IMFP to be estimated from experimental data, and the resulting values can be compared with our ML predictions. It shall be noted that Eq. (15) is an empirical formula, and this method allows us to do only a rough comparison because it has an unknown amount of error. In particular, the method in Ref. [62] is very approximate, especially for the value of x in Eq. (12). The x changes with the materials because of the related lattice relaxation and surface core level shift in materials. This means that the trend of the curve in Fig. 6 is more important than the absolute values.

Figure 6 presents the comparison of the IMFPs determined by our ML method, TPP-2M formula, and using Eq. (15). Because of the robustness of our ML method, the trend predicted by the ML approach is consistent for the seven materials, and

TABLE VII. Properties of compounds used to predict IMFPs. Note that the values of Z, M, N_v, and R are estimated using the VCA.

Element	Z	M	$\rho(g/cm^3)$	N_v	E_p (eV)	E_g (eV)	E_F (eV)	R (pm)
AgBr	41	93.89	6.48	9	22.71	2.68	6.95	137.5
AgCl	32	71.66	5.59	9	24.14	3.25	7.51	130
AgI	50	117.39	5.72	9	19.08	2.92	6.15	150
Al_2O_3	10	20.39	3.97	4.8	27.86	8.63	16.63	86
AlAs	23	50.95	3.73	4	15.59	2.16	7.43	120
AlN	10	20.49	3.26	4	22.99	6	12.03	95
AlSb	32	74.37	4.28	4	13.83	1.62	6.94	135
c-BN	6	12.41	3.49	4	30.56	7.2	15.71	75
h-BN	6	12.41	2.3	4	24.81	5	13.77	75
CdS	32	72.24	4.8	9	22.28	2.46	6.88	127.5
CdSe	41	95.69	5.66	9	21.03	1.7	6.23	135
CdTe	50	120.01	5.85	9	19.09	1.51	6.13	147.5
GaAs	32	72.32	5.32	4	15.63	1.47	8.54	122.5
GaN	19	41.86	6.09	4	21.98	3.4	10.49	97.5
GaP	23	50.35	4.13	4	16.51	2.26	8.93	115
GaSb	41	95.74	5.61	4	13.95	0.73	7.75	137.5
GaSe	32.5	74.34	5.07	4.5	15.96	1.98	9.71	122.5
InAs	41	94.87	5.67	4	14.09	0.36	6.48	135
InP	32	72.90	4.79	4	14.77	1.38	7.36	127.5
InSb	50	118.29	5.78	4	12.74	0.18	6.36	150
KBr	27	59.50	2.75	4	12.39	7.26	9.86	167.5
KCl	18	37.27	1.98	4	13.28	7.4	10.1	160
MgF_2	10	20.77	3.177	5.33	26.03	10.95	16.45	83.33
MgO	10	20.15	3.576	4	24.28	7.69	13.99	105
NaCl	14	29.22	2.165	4	15.69	9	13.1	140
$NbC_{0.712}$	26.44	59.26	7.746	4.58	22.31	0	7.4	113.81
$NbC_{0.844}$	24.98	55.88	7.769	4.54	22.9	0	7.4	110.67
$NbC_{0.93}$	24.13	53.93	7.781	4.52	23.27	0	7.4	108.86
PbS	49	119.63	7.62	5	16.26	0.42	5.63	140
PbSe	58	143.08	8.29	5	15.51	0.29	5.29	147.5
PbTe	67	167.40	8.27	5	14.32	0.32	4.86	160
SiC	10	20.05	3.22	4	23.1	2.31	9.26	90
SiO2	10	20.00	2.19	5.33	22.02	9.1	19.1	85
SnTe	51	123.16	6.47	5	14.77	0.19	8.44	142.5
$TiC_{0.7}$	15.41	33.11	4.627	4	21.54	0	5.7	111.18
$TiC_{0.95}$	14.21	30.41	4.843	4	23	0	5.7	105.90
$VC_{0.76}$	15.66	34.13	5.582	4.57	24.91	0	7.5	106.93
$VC_{0.86}$	15.14	32.94	5.605	4.54	25.32	0	7.5	104.95
$Y_3Al_5O_{12}$	13.9	29.68	4.554	4.8	24.73	6.5	13	94.25
ZnS	23	48.72	4.09	9	25.05	3.81	9.18	117.5
ZnSe	32	72.17	5.26	9	23.34	2.68	8.1	125
ZnTe	41	96.49	5.64	9	20.9	2.25	7.67	137.5

the high energy region agrees well with the results of the TPP-2M formula. Considering that our training set includes the calculated data for optical constants, instead of experimental data, the agreement between our ML data and experimental data is very valuable. For most of the transition and lanthanide metals, the GPR result is close to the experimental data, even if the comparison is only for the lower energy region because of the lack of experimental data (obviously the results for the higher energy region will be even better). The largest inaccuracy is observed for the lanthanide metals. One reason for this is that there are only three lanthanide metals available for our training dataset. A complete series of data should improve the performance for this group of elements. These results demonstrate that the ML method is very reliable, which is because it is only dependent on the reliability of the input data; it does not require any artificial or subjective factors that are often included in empirical formulas. More details of the ML prediction data used to determine the IMFPs of transition and lanthanide metals [uncolored elements in Fig. 2(a)] are described in the Appendix.

Finally, the error caused by ML and the training data must be considered. Shinotsuka *et al.* [24] calculated IMFP data using the FPA with optical constants. Because of the lack of experimental data, the optical constants are not complete, which causes inaccuracy in the training data, especially in the very low energy (<50 eV) region, which is sensitive to these constants. Together with the reason of the FPA-calculation model, it was pointed out by Shinotsuka *et al.* that the FPA-calculated IMFP data are relatively not reliable in the very low energy (<50 eV) region and very high energy (>200

TABLE VIII. RMSDs of GPR predicted compound IMFP results using VCA.

Material	RMSD (%)
AgBr	6.93
AgCl	4.87
AgI	10.76
Al_2O_3	14.20
AlAs	8.14
AlN	7.41
AlSb	6.00
c-BN	9.99
h-BN	4.45
CdS	7.32
CdSe	10.56
CdTe	12.20
GaAs	6.80
GaN	9.37
GaP	7.79
GaSb	2.05
GaSe	5.69
InAs	3.31
InP	6.85
InSb	4.99
KBr	21.35
KCl	22.62
MgF_2	24.35
MgO	9.95
NaCl	24.72
$NbC_{0.712}$	6.38
$NbC_{0.844}$	7.25
$NbC_{0.93}$	7.69
PbS	11.82
PbSe	15.14
PbTe	16.21
SiC	9.83
SiO_2	18.38
SnTe	8.42
$TiC_{0.7}$	10.59
$TiC_{0.95}$	12.86
$VC_{0.76}$	7.15
$VC_{0.86}$	7.94
$Y_3Al_5O_{12}$	4.29
ZnS	8.79
ZnSe	7.88
ZnTe	11.22
Average	10.11

keV) region. In addition, according to Ref. [65], N_v cannot be reliably evaluated in many cases, at least for the transition and lanthanide metals predicted in this paper.

III. DISCUSSION

So far, the results shown above are limited to the most preliminary predictions and tests in the elemental material IMFPs. Therefore, not limited to elemental material IMFPs, to test the extension ability from elemental materials to compounds, our initial attempt was to use the virtual crystal approximation (VCA) [66] on the expansion from elemental materials to compounds. The VCA is a well-received simple approximation used on the first-principles calculation, in which the compounds are treated as primitive-periodicity-styled crystals. However, the averaged potential of each atoms is simply used for the total potential of the "virtual" system. Similarly, a very natural idea is to apply the VCA directly on elemental IMFPs to predict compounds.

Supposing that, for a compound A_xB_y, the IMFP can be calculated through the accumulation of cross-sections of A and B as follows, using VCA:

$$\frac{1}{\lambda_{A_xB_y}} = n_{A_xB_y}\sigma_{A_xB_y} = n_A{}'\sigma_A + n_B{}'\sigma_B, \tag{16}$$

where n' is the number of following atoms in the compound in unit volume, and σ is the cross-section. Then the IMFP of A_xB_y can be derived according to the following equation:

$$\lambda_{A_xB_y} = \left(\frac{n_A{}'}{n_A\lambda_A} + \frac{n_B{}'}{n_B\lambda_B}\right)^{-1}. \tag{17}$$

However, this is very limited because the IMFPs are not always available for A or B, e.g., for the situation of oxides and halides. We tried to predict the IMFPs of SiC using the IMFPs of silicon and graphite. The predictive result is shown in Fig. 7. The calculated IMFPs of SiC [25], Si, and C (graphite) [24] based on experimental data are shown in dots for comparison. We noted that the curve of predicted values is not lying between the curves of Si and C (graphite), as usual, because the SiC IMFP prediction is a weighted harmonized average value of its components, according to Eq. (17). The predicted IMFPs of SiC do not agree with the FPA-calculated IMFPs with the RMSD of 14.16%; even the raw data of elemental graphite and silicon is closer than the prediction here, as shown in Fig. 7. Note that the energy gap of SiC (2.31 eV) is even larger than the energy gap of Si (1.1 eV) and graphite (0 eV), so the poor prediction here is probably due to the large energy gap of SiC. In fact, Table IV also indicates that the semiconductors always have very large RMSDs, for example, C (diamond) and Ge. Semiconductors,

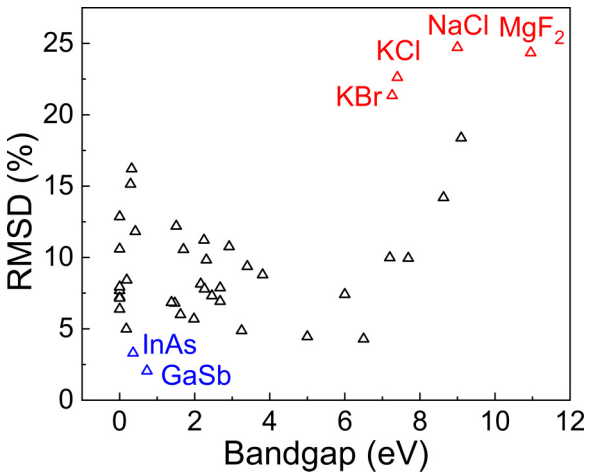

FIG. 8. The relationship between root mean square deviations (RMSDs) of Gaussian process regression (GPR)-predicted compound inelastic mean free path (IMFP) results using virtual crystal approximation and the bandgap of compounds.

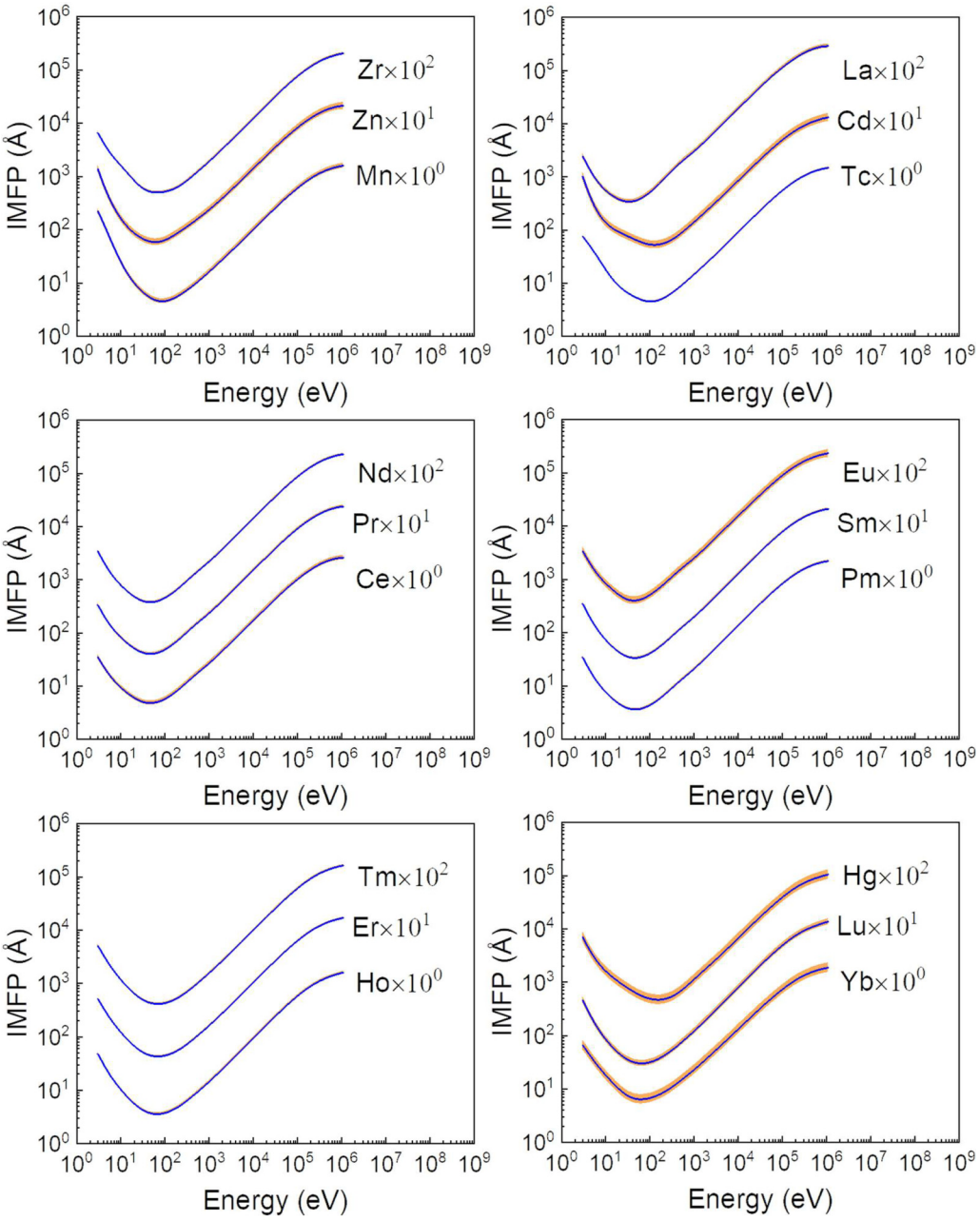

FIG. 9. Prediction result in Table IX with variance.

together with insulators, have nonzero energy gaps, different from most other materials in the training set, which led to slower learning rates and poor results, so the prediction results of semiconductors and insulators were not as accurate as those of metals. In other words, it seems undesirable using the VCA on IMFPs to predict compound IMFPs.

Alternatively, we now try to predict compound IMFPs with ML using the model trained with elemental material IMFPs. Obviously, some features of compounds should be estimated with VCA and used as the input features in ML, as shown in Table VII. With the comparison with the true values of the compound IMFP database, which is taken from Ref. [25] calculated by Shinotsuka *et.al.* for accuracy validation, the prediction RMSD results are shown in Table VIII.

As mentioned before, for a better discussion of the relationship between RMSDs and bandgap energies, Fig. 8 shows a clearly positive correlation between them. There are some materials with extra-large errors at the right side of Fig. 8, e.g., KBr, KCl, MgF_2, and NaCl with red color. These materials are all halides with large bandgap energies, in which the atomics are combined with ionic bonds. In these compounds, the valence electrons are largely biased to halogen atoms in the compounds, so the physical and chemical properties of these materials are very different from the elemental materials included in our training set in its formation. Therefore, the predicted IMFPs of these compounds from the IMFPs of elemental materials have larger deviations. Meanwhile, there are also some materials with good predictions, such as GaSb

TABLE IX. Prediction result of transition and lanthanide metals.

Energy	Inelastic mean free path (Å)																	
(eV)	Mn	Zn	Zr	Tc	Cd	La	Ce	Pr	Nd	Pm	Sm	Eu	Ho	Er	Tm	Yb	Lu	Hg
3.0	220.35	138.68	65.95	76.00	101.88	24.30	34.86	33.47	34.22	34.35	34.82	34.16	47.87	51.25	50.79	65.78	45.84	70.19
3.3	187.99	110.89	57.29	68.60	80.82	20.83	30.61	29.03	29.56	29.56	29.84	29.69	41.35	44.92	44.36	58.97	38.91	59.85
3.7	153.86	85.59	48.52	60.50	62.00	17.38	26.25	24.56	24.88	24.77	24.87	25.15	34.78	38.40	37.76	51.61	32.09	49.83
4.1	127.60	68.56	42.02	53.90	49.58	14.86	22.96	21.24	21.42	21.25	21.23	21.77	29.91	33.44	32.77	45.78	27.14	42.66
4.5	107.21	56.61	37.10	48.40	41.01	12.97	20.43	18.73	18.80	18.59	18.49	19.21	26.20	29.60	28.91	41.11	23.43	37.34
5.0	87.73	46.11	32.45	42.70	33.62	11.21	18.02	16.36	16.34	16.09	15.92	16.78	22.69	25.88	25.21	36.48	19.98	32.43
5.5	73.08	38.72	28.96	38.02	28.52	9.91	16.18	14.57	14.49	14.22	14.00	14.96	20.03	23.03	22.36	32.82	17.42	28.79
6.0	61.86	33.30	26.26	34.11	24.84	8.93	14.74	13.20	13.06	12.78	12.53	13.56	17.96	20.77	20.13	29.88	15.46	26.00
6.7	50.19	27.85	23.36	29.66	21.21	7.89	13.19	11.73	11.54	11.26	10.97	12.06	15.74	18.32	17.70	26.62	13.38	23.06
7.4	41.73	23.97	21.14	26.11	18.67	7.13	12.02	10.63	10.40	10.11	9.81	10.92	14.05	16.44	15.85	24.07	11.84	20.86
8.2	34.67	20.75	19.15	22.89	16.61	6.48	10.98	9.66	9.40	9.12	8.80	9.93	12.56	14.76	14.20	21.75	10.50	18.94
9.0	29.49	18.37	17.55	20.33	15.10	5.99	10.17	8.91	8.63	8.35	8.01	9.15	11.40	13.43	12.91	19.89	9.46	17.45
10.0	24.75	16.16	15.92	17.83	13.73	5.52	9.36	8.17	7.87	7.59	7.25	8.38	10.25	12.12	11.63	18.01	8.46	16.00
11.0	21.29	14.51	14.59	15.88	12.72	5.15	8.72	7.58	7.27	7.00	6.66	7.76	9.35	11.08	10.62	16.50	7.68	14.86
12.2	18.25	13.02	13.26	14.09	11.82	4.81	8.11	7.02	6.70	6.44	6.10	7.17	8.49	10.08	9.65	15.03	6.95	13.77
13.5	15.85	11.81	12.06	12.60	11.08	4.53	7.58	6.54	6.22	5.97	5.63	6.65	7.74	9.22	8.82	13.74	6.34	12.83
14.9	13.94	10.81	11.00	11.39	10.48	4.29	7.11	6.12	5.80	5.56	5.22	6.20	7.11	8.47	8.11	12.61	5.81	12.01
16.4	12.41	9.99	10.06	10.39	9.98	4.09	6.71	5.75	5.44	5.21	4.87	5.80	6.56	7.83	7.50	11.63	5.36	11.29
18.2	11.03	9.23	9.14	9.47	9.50	3.90	6.33	5.40	5.09	4.87	4.55	5.42	6.03	7.22	6.92	10.69	4.94	10.58
20.1	9.94	8.61	8.36	8.74	9.10	3.76	6.00	5.11	4.81	4.60	4.28	5.10	5.59	6.71	6.43	9.90	4.59	9.96
22.2	9.02	8.08	7.68	8.12	8.73	3.64	5.72	4.85	4.56	4.36	4.05	4.82	5.21	6.26	6.00	9.21	4.29	9.39
24.5	8.24	7.63	7.09	7.59	8.39	3.54	5.48	4.64	4.35	4.16	3.86	4.59	4.88	5.88	5.63	8.61	4.03	8.87
27.1	7.57	7.23	6.59	7.13	8.07	3.47	5.28	4.45	4.18	3.99	3.70	4.39	4.60	5.54	5.31	8.09	3.80	8.38
30.0	6.98	6.89	6.17	6.73	7.76	3.43	5.11	4.30	4.03	3.85	3.57	4.24	4.35	5.25	5.03	7.65	3.61	7.93
33.1	6.50	6.62	5.84	6.39	7.48	3.41	4.98	4.19	3.93	3.75	3.47	4.13	4.15	5.01	4.81	7.30	3.45	7.52
36.6	6.07	6.38	5.57	6.07	7.20	3.42	4.89	4.10	3.85	3.68	3.40	4.05	3.98	4.81	4.62	7.02	3.32	7.14
40.4	5.71	6.19	5.37	5.80	6.93	3.46	4.83	4.05	3.80	3.63	3.36	4.01	3.85	4.65	4.46	6.80	3.22	6.79
44.7	5.41	6.04	5.22	5.55	6.68	3.52	4.81	4.03	3.78	3.61	3.35	4.01	3.74	4.52	4.34	6.63	3.14	6.47
49.4	5.15	5.94	5.11	5.33	6.44	3.61	4.81	4.04	3.79	3.62	3.35	4.04	3.67	4.42	4.25	6.51	3.08	6.17
54.6	4.94	5.88	5.05	5.14	6.21	3.73	4.85	4.07	3.82	3.65	3.39	4.10	3.62	4.35	4.18	6.44	3.04	5.91
60.3	4.78	5.86	5.02	4.97	6.01	3.87	4.92	4.13	3.88	3.70	3.44	4.19	3.59	4.31	4.14	6.42	3.02	5.68
66.7	4.66	5.88	5.01	4.82	5.82	4.05	5.02	4.22	3.96	3.78	3.52	4.31	3.58	4.30	4.13	6.43	3.02	5.47
73.7	4.59	5.95	5.03	4.71	5.66	4.26	5.15	4.33	4.07	3.88	3.62	4.47	3.60	4.30	4.13	6.47	3.03	5.28
81.5	4.55	6.05	5.07	4.62	5.53	4.50	5.31	4.48	4.21	4.01	3.75	4.66	3.63	4.33	4.16	6.55	3.06	5.13
90.0	4.54	6.21	5.13	4.56	5.42	4.78	5.50	4.65	4.37	4.16	3.90	4.87	3.69	4.38	4.21	6.65	3.11	4.99
99.5	4.58	6.40	5.20	4.52	5.34	5.11	5.73	4.86	4.56	4.34	4.07	5.12	3.76	4.46	4.27	6.79	3.17	4.88
109.9	4.65	6.64	5.30	4.52	5.29	5.47	5.99	5.09	4.78	4.55	4.27	5.40	3.86	4.55	4.36	6.95	3.25	4.80
121.5	4.75	6.92	5.41	4.56	5.26	5.89	6.30	5.37	5.03	4.79	4.50	5.72	3.98	4.67	4.48	7.14	3.34	4.74
134.3	4.89	7.24	5.56	4.62	5.27	6.36	6.65	5.68	5.32	5.06	4.76	6.09	4.11	4.81	4.61	7.36	3.46	4.70
148.4	5.06	7.60	5.73	4.72	5.30	6.89	7.04	6.03	5.65	5.37	5.06	6.49	4.28	4.98	4.78	7.62	3.59	4.68
164.0	5.26	7.99	5.94	4.85	5.36	7.48	7.49	6.43	6.02	5.72	5.39	6.94	4.46	5.18	4.96	7.90	3.74	4.68

TABLE IX. (*Continued*).

Energy	Inelastic mean free path (Å)																	
(eV)	Mn	Zn	Zr	Tc	Cd	La	Ce	Pr	Nd	Pm	Sm	Eu	Ho	Er	Tm	Yb	Lu	Hg
181.3	5.50	8.43	6.18	5.01	5.46	8.14	7.99	6.88	6.44	6.11	5.75	7.44	4.67	5.41	5.18	8.23	3.92	4.71
200.3	5.76	8.90	6.47	5.21	5.59	8.86	8.55	7.37	6.89	6.55	6.16	7.99	4.91	5.66	5.43	8.60	4.11	4.77
221.4	6.06	9.40	6.81	5.44	5.75	9.67	9.17	7.92	7.40	7.03	6.62	8.60	5.19	5.96	5.70	9.02	4.33	4.85
244.7	6.39	9.95	7.19	5.71	5.95	10.54	9.85	8.53	7.96	7.55	7.11	9.26	5.49	6.29	6.02	9.48	4.58	4.96
270.4	6.75	10.53	7.63	6.02	6.19	11.49	10.59	9.19	8.57	8.13	7.65	9.98	5.82	6.65	6.37	9.99	4.86	5.11
298.9	7.15	11.15	8.13	6.37	6.46	12.51	11.40	9.91	9.24	8.76	8.24	10.76	6.20	7.06	6.76	10.56	5.16	5.28
330.3	7.58	11.81	8.68	6.75	6.78	13.59	12.27	10.68	9.95	9.43	8.88	11.59	6.60	7.51	7.19	11.18	5.50	5.50
365.0	8.06	12.53	9.29	7.18	7.15	14.73	13.21	11.50	10.72	10.16	9.56	12.47	7.05	8.00	7.66	11.87	5.86	5.76
403.4	8.59	13.29	9.96	7.66	7.57	15.93	14.21	12.38	11.53	10.94	10.29	13.41	7.54	8.54	8.18	12.61	6.27	6.06
445.9	9.17	14.12	10.68	8.19	8.04	17.19	15.27	13.32	12.41	11.77	11.07	14.41	8.08	9.14	8.75	13.42	6.71	6.40
492.7	9.80	15.01	11.47	8.76	8.56	18.50	16.40	14.31	13.34	12.65	11.90	15.46	8.66	9.78	9.37	14.30	7.20	6.80
544.6	10.49	15.97	12.32	9.39	9.15	19.87	17.60	15.37	14.32	13.59	12.78	16.57	9.29	10.49	10.04	15.25	7.72	7.25
601.8	11.25	17.01	13.24	10.08	9.79	21.31	18.88	16.49	15.38	14.59	13.72	17.75	9.98	11.25	10.78	16.28	8.30	7.76
665.1	12.07	18.14	14.23	10.83	10.50	22.83	20.25	17.69	16.50	15.66	14.74	19.00	10.72	12.08	11.58	17.40	8.92	8.33
735.1	12.98	19.36	15.30	11.65	11.28	24.44	21.71	18.98	17.71	16.82	15.83	20.34	11.53	12.99	12.44	18.60	9.61	8.96
812.4	13.96	20.68	16.47	12.54	12.14	26.17	23.29	20.37	19.02	18.06	17.00	21.79	12.41	13.97	13.39	19.91	10.35	9.65
897.8	15.03	22.12	17.73	13.50	13.06	28.03	25.01	21.88	20.43	19.41	18.28	23.36	13.37	15.04	14.41	21.33	11.16	10.40
992.3	16.20	23.67	19.11	14.55	14.06	30.06	26.87	23.52	21.98	20.88	19.67	25.07	14.42	16.20	15.53	22.87	12.04	11.22
1096.6	17.46	25.36	20.61	15.69	15.15	32.29	28.90	25.31	23.66	22.49	21.19	26.95	15.55	17.46	16.74	24.55	13.00	12.11
1212.0	18.84	27.20	22.26	16.92	16.33	34.74	31.13	27.28	25.51	24.26	22.85	29.01	16.80	18.84	18.07	26.37	14.04	13.06
1339.4	20.33	29.21	24.05	18.25	17.60	37.45	33.57	29.44	27.54	26.19	24.68	31.27	18.16	20.34	19.51	28.35	15.18	14.09
1480.3	21.94	31.39	26.02	19.70	18.97	40.44	36.25	31.81	29.78	28.32	26.70	33.77	19.64	21.97	21.08	30.51	16.43	15.19
1636.0	23.69	33.77	28.16	21.26	20.45	43.75	39.20	34.43	32.23	30.66	28.91	36.51	21.26	23.75	22.79	32.85	17.79	16.37
1808.0	25.59	36.38	30.50	22.96	22.06	47.40	42.43	37.30	34.93	33.23	31.34	39.52	23.02	25.69	24.65	35.40	19.27	17.64
1998.2	27.66	39.23	33.04	24.80	23.80	51.43	45.97	40.45	37.88	36.05	34.00	42.83	24.95	27.80	26.68	38.16	20.88	19.00
2208.3	29.90	42.35	35.81	26.81	25.69	55.85	49.84	43.90	41.12	39.14	36.92	46.44	27.05	30.11	28.89	41.16	22.64	20.48
2440.6	32.34	45.75	38.81	28.99	27.74	60.69	54.06	47.66	44.66	42.51	40.11	50.37	29.34	32.61	31.30	44.41	24.56	22.07
2697.3	35.00	49.48	42.08	31.36	29.96	65.97	58.67	51.77	48.51	46.19	43.59	54.65	31.84	35.34	33.93	47.92	26.65	23.79
2981.0	37.89	53.54	45.63	33.95	32.38	71.71	63.67	56.23	52.71	50.20	47.38	59.29	34.55	38.31	36.79	51.72	28.93	25.65
3294.5	41.04	57.96	49.49	36.77	35.00	77.95	69.11	61.08	57.28	54.56	51.49	64.31	37.51	41.55	39.90	55.84	31.41	27.69
3641.0	44.47	62.78	53.69	39.84	37.85	84.70	75.00	66.34	62.23	59.29	55.97	69.73	40.72	45.06	43.28	60.29	34.11	29.90
4023.9	48.21	68.01	58.25	43.20	40.95	91.99	81.40	72.04	67.60	64.42	60.82	75.59	44.22	48.89	46.96	65.10	37.05	32.31
4447.1	52.28	73.70	63.23	46.85	44.32	99.87	88.32	78.22	73.42	69.99	66.08	81.90	48.03	53.06	50.98	70.31	40.25	34.94
4914.8	56.72	79.86	68.66	50.84	47.98	108.39	95.82	84.91	79.73	76.03	71.78	88.73	52.17	57.59	55.34	75.95	43.74	37.81
5431.7	61.54	86.55	74.58	55.18	51.96	117.57	103.94	92.15	86.57	82.57	77.97	96.09	56.68	62.52	60.09	82.05	47.54	40.95
6002.9	66.79	93.80	81.04	59.91	56.28	127.50	112.75	100.00	93.98	89.67	84.67	104.05	61.58	67.89	65.27	88.67	51.67	44.37
6634.2	72.49	101.66	88.07	65.05	60.97	138.24	122.30	108.52	102.02	97.37	91.95	112.66	66.92	73.74	70.90	95.84	56.17	48.10
7332.0	78.68	110.18	95.74	70.65	66.07	149.86	132.66	117.75	110.75	105.73	99.86	121.98	72.73	80.10	77.03	103.61	61.06	52.17
8103.1	85.41	119.42	104.08	76.74	71.62	162.44	143.91	127.77	120.23	114.81	108.44	132.07	79.05	87.02	83.70	112.04	66.40	56.60
8955.3	92.72	129.46	113.16	83.36	77.64	176.09	156.13	138.65	130.53	124.68	117.77	143.01	85.93	94.55	90.96	121.18	72.20	61.42
9897.1	100.65	140.35	123.02	90.55	84.17	190.90	169.39	150.47	141.71	135.40	127.91	154.87	93.40	102.74	98.86	131.07	78.51	66.65

TABLE IX. (*Continued*).

Energy	Inelastic mean free path (Å)																	
(eV)	Mn	Zn	Zr	Tc	Cd	La	Ce	Pr	Nd	Pm	Sm	Eu	Ho	Er	Tm	Yb	Lu	Hg
10938.0	109.26	152.17	133.72	98.36	91.27	206.99	183.80	163.30	153.87	147.04	138.92	167.73	101.53	111.63	107.44	141.78	85.37	72.34
12088.4	118.59	165.01	145.33	106.84	98.96	224.45	199.45	177.24	167.07	159.70	150.89	181.68	110.36	121.30	116.76	153.38	92.82	78.49
13359.7	128.71	178.92	157.91	116.04	107.30	243.41	216.44	192.38	181.41	173.45	163.90	196.79	119.95	131.79	126.88	165.92	100.92	85.15
14764.8	139.68	194.01	171.56	126.01	116.33	264.00	234.88	208.80	196.98	188.38	178.01	213.16	130.35	143.17	137.86	179.46	109.70	92.35
16317.6	151.55	210.35	186.34	136.83	126.10	286.35	254.88	226.63	213.87	204.58	193.33	230.88	141.63	155.50	149.75	194.08	119.22	100.13
18033.7	164.41	228.03	202.35	148.53	136.64	310.57	276.54	245.94	232.18	222.15	209.94	250.05	153.85	168.85	162.63	209.84	129.53	108.52
19930.4	178.32	247.13	219.69	161.21	148.03	336.81	300.00	266.86	252.02	241.18	227.94	270.77	167.09	183.29	176.57	226.82	140.70	117.55
22026.5	193.35	267.74	238.46	174.91	160.31	365.22	325.37	289.50	273.49	261.78	247.42	293.14	181.39	198.90	191.63	245.10	152.77	127.27
24343.0	209.58	289.96	258.78	189.71	173.54	395.91	352.78	313.96	296.70	284.05	268.47	317.28	196.86	215.75	207.90	264.76	165.81	137.74
26903.2	227.10	313.88	280.74	205.69	187.78	429.04	382.36	340.38	321.76	308.11	291.22	343.28	213.55	233.93	225.45	285.88	179.90	148.99
29732.6	245.96	339.61	304.47	222.93	203.09	464.73	414.23	368.85	348.79	334.06	315.75	371.26	231.55	253.52	244.37	308.54	195.09	161.08
32859.6	266.26	367.23	330.07	241.48	219.55	503.12	448.51	399.49	377.89	362.01	342.17	401.33	250.95	274.62	264.73	332.84	211.46	174.06
36315.5	288.07	396.87	357.63	261.44	237.21	544.31	485.31	432.43	409.18	392.07	370.59	433.58	271.83	297.29	286.62	358.84	229.07	188.01
40134.8	311.46	428.61	387.24	282.87	256.15	588.45	524.77	467.76	442.74	424.32	401.09	468.12	294.26	321.63	310.13	386.63	248.01	202.96
44355.9	336.50	462.55	419.00	305.84	276.43	635.61	566.97	505.57	478.69	458.87	433.76	505.03	318.32	347.73	335.34	416.30	268.33	218.98
49020.8	363.26	498.80	452.96	330.42	298.10	685.88	612.01	545.95	517.08	495.79	468.69	544.40	344.07	375.64	362.31	447.91	290.09	236.11
54176.4	391.79	537.42	489.19	356.64	321.20	739.35	659.96	588.97	558.02	535.16	505.95	586.27	371.59	405.43	391.10	481.50	313.36	254.42
59874.1	422.12	578.48	527.74	384.56	345.79	796.07	710.89	634.67	601.53	577.03	545.58	630.71	400.91	437.16	421.78	517.13	338.17	273.92
66171.2	454.31	622.03	568.62	414.21	371.86	856.05	764.82	683.09	647.68	621.43	587.62	677.74	432.08	470.88	454.37	554.82	364.54	294.66
73130.4	488.37	668.10	611.87	445.59	399.44	919.33	821.81	734.28	696.45	668.39	632.09	727.34	465.09	506.58	488.91	594.58	392.52	316.64
80821.6	524.31	716.69	657.49	478.74	428.51	985.89	881.80	788.19	747.86	717.89	678.98	779.51	499.98	544.30	525.40	636.39	422.09	339.85
89321.7	562.13	767.77	705.46	513.61	459.05	1055.75	944.84	844.81	801.88	769.93	728.28	834.20	536.70	584.01	563.82	680.24	453.22	364.28
98715.8	601.81	821.32	755.74	550.17	491.00	1128.83	1010.86	904.11	858.48	824.44	779.92	891.33	575.20	625.68	604.13	726.04	485.88	389.90
109097.8	643.28	877.24	808.30	588.40	524.30	1205.09	1079.79	965.98	917.55	881.37	833.85	950.82	615.43	669.21	646.28	773.71	520.00	416.63
120571.7	686.47	935.45	863.02	628.19	558.87	1284.43	1151.54	1030.34	979.02	940.61	889.94	1012.49	657.28	714.55	690.16	823.17	555.48	444.39

TABLE IX. (*Continued*).

Energy	Inelastic mean free path (Å)																	
(eV)	Mn	Zn	Zr	Tc	Cd	La	Ce	Pr	Nd	Pm	Sm	Eu	Ho	Er	Tm	Yb	Lu	Hg
133252.4	731.29	995.80	919.84	669.45	594.61	1366.72	1225.97	1097.03	1042.73	1002.01	948.07	1076.22	700.62	761.55	735.65	874.24	592.23	473.09
147266.6	777.59	1058.11	978.52	712.08	631.38	1451.81	1302.93	1165.92	1108.54	1065.42	1008.07	1141.75	745.28	810.05	782.60	926.72	630.06	502.60
162754.8	825.18	1122.15	1038.92	755.91	669.07	1539.43	1382.20	1236.75	1176.17	1130.63	1069.70	1208.87	791.10	859.84	830.79	980.42	668.81	532.76
179871.9	873.82	1187.68	1100.75	800.72	707.49	1629.33	1463.42	1309.24	1245.43	1197.35	1132.71	1277.29	837.80	910.71	880.01	1035.07	708.29	563.42
198789.2	923.25	1254.35	1163.70	846.31	746.45	1721.08	1546.32	1383.06	1315.92	1265.29	1196.80	1346.64	885.18	962.37	929.99	1090.39	748.24	594.39
219696.0	973.13	1321.84	1227.35	892.42	785.76	1814.22	1630.38	1457.84	1387.30	1334.07	1261.60	1416.61	932.93	1014.47	980.39	1146.04	788.44	625.48
242801.6	1023.13	1389.73	1291.34	938.73	825.18	1908.18	1715.14	1533.13	1459.12	1403.27	1326.72	1486.76	980.75	1066.72	1030.91	1201.71	828.63	656.49
268337.3	1072.85	1457.54	1355.19	984.94	864.49	2002.30	1799.95	1608.42	1530.95	1472.41	1391.78	1556.68	1028.35	1118.74	1081.21	1257.04	868.58	687.26
296558.6	1121.94	1524.79	1418.44	1030.67	903.46	2095.90	1884.21	1683.18	1602.25	1541.06	1456.30	1626.00	1075.45	1170.20	1130.94	1311.69	908.07	717.58
327747.9	1170.01	1591.07	1480.64	1075.62	941.85	2188.11	1967.21	1756.87	1672.48	1608.72	1519.88	1694.34	1121.81	1220.73	1179.78	1365.37	946.91	747.31
362217.4	1216.75	1655.81	1541.42	1119.49	979.47	2278.14	2048.24	1828.94	1741.21	1674.90	1582.12	1761.29	1167.21	1270.11	1227.47	1417.75	985.01	776.37
400312.2	1261.89	1718.58	1600.33	1161.96	1016.16	2365.10	2126.52	1898.89	1807.92	1739.16	1642.67	1826.59	1211.55	1318.01	1273.80	1468.69	1022.26	804.67
442413.4	1305.18	1778.89	1657.18	1202.82	1051.82	2448.16	2201.41	1966.21	1872.19	1801.15	1701.26	1889.99	1254.70	1364.36	1318.62	1517.99	1058.72	832.20
488942.4	1346.54	1836.33	1711.71	1241.91	1086.33	2526.50	2272.22	2030.44	1933.66	1860.46	1757.56	1951.24	1296.70	1409.00	1361.85	1565.56	1094.46	858.99
540364.9	1385.89	1890.47	1763.88	1279.09	1119.70	2599.32	2338.30	2091.27	1991.96	1916.86	1811.42	2010.20	1337.58	1451.88	1403.49	1611.28	1129.64	885.10
597195.6	1423.21	1940.84	1813.55	1314.32	1151.86	2665.75	2399.00	2148.18	2046.82	1970.02	1862.69	2066.67	1377.46	1493.04	1443.51	1655.08	1164.39	910.60
660003.2	1458.45	1986.92	1860.72	1347.57	1182.85	2724.96	2453.81	2200.85	2097.83	2019.67	1911.13	2120.36	1416.41	1532.39	1481.94	1696.88	1198.92	935.53
729416.4	1491.59	2028.15	1905.24	1378.76	1212.61	2776.13	2501.90	2248.74	2144.67	2065.48	1956.50	2170.95	1454.45	1569.97	1518.76	1736.48	1233.36	959.98
806129.8	1522.44	2063.81	1946.93	1407.80	1240.97	2818.19	2542.55	2291.34	2186.70	2106.93	1998.39	2217.84	1491.49	1605.54	1553.82	1773.54	1267.71	983.83
890911.2	1550.71	2092.96	1985.41	1434.50	1267.74	2849.97	2574.78	2327.77	2223.31	2143.43	2036.20	2260.22	1527.29	1638.74	1586.83	1807.47	1301.82	1006.93
984609.1	1575.98	2114.51	2019.99	1458.48	1292.50	2870.19	2597.53	2357.06	2253.51	2174.00	2069.09	2297.00	1561.35	1669.13	1617.26	1837.60	1335.37	1029.01
1088161.4	1597.42	2127.11	2049.75	1479.18	1314.65	2877.27	2609.34	2377.83	2276.09	2197.51	2095.90	2326.64	1592.91	1695.78	1644.37	1862.73	1367.73	1049.54

and InAs with blue color at the left side of Fig. 8. These materials show strong metallic characteristics with relatively small bandgaps, in which the valence electrons show low bias between the atomics and can be predicted by the VCA of its formation elements. Moreover, the materials with small bandgaps are also better predicted in our model, which was trained by elemental materials. In the prediction of small bandgap compounds, these similarly well-predicted small bandgap elemental materials can have larger contribution due to their similar electron properties. Although for most of the compounds, the electron behaviors are very different from their component elements, leading to poor prediction, this ML model holds the potential to be used for predicting the IMFPs of alloys as a mixture of multiple elemental materials, whose electric properties are often similar to that of its components. Thus, the IMFPs of alloys will be easily predicted with only elemental materials.

IV. CONCLUSIONS

Based on the existing IMFP database, we developed an ML technique to determine IMFPs from simple material properties. The obtained ML model achieved a robust description of IMFPs over a wide energy range, overcoming the limitation of the TPP-2M formula in the very low energy (<50 eV) range. In the LOOCV testing, the ML model showed reliable performance in IMFP prediction. Based on the developed ML model, we predicted IMFPs for several transition and lanthanide metals that were not included in the existing database because of missing optical constants. Improved predictions of IMFPs were achieved by our ML method, proving its superiority to traditional empirical formula fitting methods. This paper is only an initial example of using ML to complete the missing part of the IMFP database, and we will extend this method to the IMFP estimation of compounds and other materials in future work. Since the GPR method is more adept at using local information for prediction, we are going to use the GPR method to predict the IMFP of metal alloys. Importantly, this ML method is not limited to IMFP prediction; it can easily be extended to any other field to determine a small number of missing data values in a specified database.

V. DATA AVAILABILITY

All data generated and/or analyzed during this study are included in this paper.

ACKNOWLEDGMENTS

This paper was supported by the National Key Research and Development Project (No. 2019YFF0216404), Chinese Education Ministry through the "111" Project 2.0 (No. BP0719016), and Grants for Basic Science Research Projects from The Sumitomo Foundation. The calculations in this paper were performed on Numerical Materials Simulator at NIMS. We also thank the Supercomputing Center of USTC for support with parallel computing. Y.S. acknowledges support from National Science Foundation awards EAR-1918134 and EAR-1918126.

The authors declare no competing financial or nonfinancial interests.

APPENDIX: GPR PREDICTION RESULT OF TRANSITION AND LANTHANIDE METALS

Table IX listed the IMFPs result of transition and lanthanide metals, predicted by our trained machine learning model. The results are also shown in Fig. 9 in blue curves, together with their variances (orange).

[1] *ISO18115 Surface Chemical Analysis—Vocabulary—Part 1: General terms and terms used in spectroscopy* (International Organisation for Standardisation, Geneva, 2010).

[2] J. D. Bourke and C. T. Chantler, Momentum-dependent lifetime broadening of electron energy loss spectra: a self-consistent coupled-plasmon model, J. Phys. Chem. Lett. **6**, 314 (2015).

[3] C. T. Chantler and J. D. Bourke, X-ray spectroscopic measurement of photoelectron inelastic mean free paths in molybdenum, J. Phys. Chem. Lett. **1**, 2422 (2010).

[4] C. J. Powell and A. Jablonski, Surface sensitivity of x-ray photoelectron spectroscopy, Nucl. Instrum. Meth. Phys. Res. A **601**, 54 (2009).

[5] W. S. M. Werner, W. Smekal, H. Stori, H. Winter, G. Stefani, A. Ruocco, F. Offi, R. Gotter, A. Morgante, and F. Tommasini, Emission-Depth-Selective Auger Photoelectron Coincidence Spectroscopy, Phys. Rev. Lett. **94**, 038302 (2005).

[6] Z. J. Ding, K. Salma, H. M. Li, Z. M. Zhang, K. Tokesi, D. Varga, J. Toth, K. Goto, and R. Shimizu, Monte Carlo simulation study of electron interaction with solids and surfaces, Surf. Interface Anal. **38**, 657 (2006).

[7] N. Cao, B. Da, Y. Ming, S. F. Mao, K. Goto, and Z. J. Ding, Monte Carlo simulation of full energy spectrum of electrons emitted from silicon in Auger electron spectroscopy, Surf. Interface Anal. **47**, 113 (2015).

[8] B. Da, Z. Y. Li, H. C. Chang, S. F. Mao, and Z. J. Ding, A Monte Carlo study of reflection electron energy loss spectroscopy spectrum of a carbon contaminated surface, J. Appl. Phys. **116**, 124307 (2014).

[9] B. Da, S. F. Mao, G. H. Zhang, X. P. Wang, and Z. J. Ding, Monte Carlo modeling of surface excitation in reflection electron energy loss spectroscopy spectrum for rough surfaces, J. Appl. Phys. **112**, 034310 (2012).

[10] Z. J. Ding and R. Shimizu, A Monte Carlo modeling of electron interaction with solids including cascade secondary electron production, Scanning **18**, 92 (1996).

[11] Y. B. Zou, S. F. Mao, B. Da, and Z. J. Ding, Surface sensitivity of secondary electrons emitted from amorphous solids: Calculation of mean escape depth by a Monte Carlo method, J. Appl. Phys. **120**, 235102 (2016).

[12] Z. J. Ding, W. S. Tan, and Y. G. Li, Improved calculation of the backscattering factor for quantitative analysis by Auger electron spectroscopy, J. Appl. Phys. **99**, 084903 (2006).

[13] R. G. Zeng, Z. J. Ding, Y. G. Li, and S. F. Mao, A calculation of backscattering factor database for quantitative analysis

by Auger electron spectroscopy, J. Appl. Phys. **104**, 114909 (2008).
[14] B. Da, K. Salma, H. Ji, S. F. Mao, G. H. Zhang, X. P. Wang, and Z. J. Ding, Surface excitation parameter for rough surfaces, Appl. Surf. Sci. **356**, 142 (2015).
[15] B. Da, Y. Sun, S. F. Mao, and Z. J. Ding, Systematic calculation of the surface excitation parameters for 22 materials, Surf. Interface Anal. **45**, 773 (2013).
[16] Z. Zheng, B. Da, S. F. Mao, and Z. J. Ding, Calculation of surface excitation parameters by a Monte Carlo method, Chin. J. Chem. Phys. **30**, 83 (2017).
[17] D. R. Penn, Electron mean-free-path calculations using a model dielectric function, Phys. Rev. B **35**, 482 (1987).
[18] N. D. Mermin, Lindhard dielectric function in the relaxation-time approximation, Phys. Rev. B **1**, 2362 (1970).
[19] B. Da, H. Shinotsuka, H. Yoshikawa, Z. J. Ding, and S. Tanuma, Extended Mermin Method for Calculating the Electron Inelastic Mean Free Path, Phys. Rev. Lett. **113**, 063201 (2014).
[20] S. Tanuma, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths for 31 materials, Surf. Interf. Anal. **11**, 577 (1988).
[21] S. Tanuma, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths. II. Data for 27 elements over the 50–2000 eV range, Surf. Interface Anal. **17**, 911 (1991).
[22] S. Tanuma, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths. III. Data for 15 inorganic compounds over the 50–2000 eV range, Surf. Interface Anal. **17**, 927 (1991).
[23] S. Tanuma, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths. V. Data for 14 organic compounds over the 50–2000 eV range, Surf. Interface Anal. **21**, 165 (1994).
[24] H. Shinotsuka, S. Tanuma, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths. X. Data for 41 elemental solids over the 50 eV to 200 keV range with the relativistic full Penn algorithm, Surf. Interface Anal. **47**, 871 (2015); **47**, 1132 (2015).
[25] H. Shinotsuka, B. Da, S. Tanuma, H. Yoshikawa, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths. XII. Data for 42 inorganic compounds over the 50 eV to 200 keV range with the full Penn algorithm, Surf. Interface Anal. **51**, 427 (2019).
[26] I. Abril, R. Garcia-Molina, C. D. Denton, F. J. Pérez-Pérez, and N. R. Arista, Dielectric description of wakes and stopping powers in solids, Phys. Rev. A **58**, 357 (1998).
[27] Y. Sun, H. Xu, B. Da, S. F. Mao, and Z. J. Ding, Calculations of energy-loss function for 26 materials, Chin. J. Chem. Phys. **29**, 663 (2016).
[28] B. Da, S. F. Mao, Y. Sun, and Z. J. Ding, A new analytical method in surface electron spectroscopy: reverse Monte Carlo method, e-J. Surf. Sci. Nanotech. **10**, 441 (2012).
[29] B. Da, Y. Sun, S. F. Mao, Z. M. Zhang, H. Jin, H. Yoshikawa, and S. Tanuma, and Z. J. Ding, A reverse Monte Carlo method for deriving optical constants of solids from reflection electron energy-loss spectroscopy spectra, J. Appl. Phys. **113**, 214303 (2013).
[30] H. Xu, B. Da, J. Tóth, K. Tőkési, and Z. J. Ding, Absolute determination of optical constants by reflection electron energy loss spectroscopy, Phys. Rev. B **95**, 195417 (2017).
[31] F. Yubero and S. Tougaard, Model for quantitative analysis of reflection-electron-energy-loss spectra. Phys. Rev. B **46**, 2486 (1992).
[32] W. S. M. Werner, Simple algorithm for quantitative analysis of reflection electron energy loss spectra (REELS), Surf. Sci. **604**, 290 (2010).
[33] H. Xu, L. H. Yang, J. Tóth, K. Tőkési, B. Da, and Z. J. Ding, Absolute determination of optical constants of three transition metals using reflection electron energy loss spectroscopy, J. Appl. Phys. **123**, 043306 (2018).
[34] L. H. Yang, M. Menyhard, A. Sulyok, K. Tőkési, and Z. J. Ding, Optical properties and excitation energies of iridium derived from reflection electron energy loss spectroscopy spectra, Appl. Surf. Sci. **456**, 999 (2018).
[35] L. H. Yang, J. Tóth, K. Tőkési, B. Da, and Z. J. Ding, Calculation of electron inelastic mean free path of three transition metals from reflection electron energy loss spectroscopy spectrum measurement data, Eur. Phys. J. D **73**, 21 (2019).
[36] W. S. M. Werner, Analysis of reflection electron energy loss spectra (REELS) for determination of the dielectric function of solids: Fe, Co, Ni, Surf. Sci. **601**, 2125 (2007).
[37] H. Jin, H. Shinotsuka, H. Yoshikawa, H. Iwai, S. Tanuma, and S. Tougaard, Measurement of optical constants of Si and SiO_2 from reflection electron energy loss spectra using factor analysis method, J. Appl. Phys. **107**, 083709 (2010).
[38] M. P. Seah and W. A. Dench, Quantitative electron spectroscopy of surfaces: a standard data base for electron inelastic mean free paths in solids, Surf. Interf. Anal. **1**, 2 (1979).
[39] A. Jablonski and C. J. Powell, Relationships between electron inelastic mean free paths, effective attenuation lengths, and mean escape depths, J. Electron. Spectros. Relat. Phenomena. **100**, 137 (1999).
[40] C. J. Powell, The quest for universal curves to describe the surface sensitivity of electron spectroscopies, J. Electron Spectrosc. Relat. Phenom. **47**, 197 (1988).
[41] H. Bethe, Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie, Ann. Phys. **397**, 325 (1930).
[42] X. Liu, Z. F. Hou, D. B. Lu, B. Da, H. Yoshikawa, S. Tanuma, Y. Sun, and Z. J. Ding, Unveiling the principle descriptor for predicting the electron inelastic mean free path based on a machine learning framework, Sci. Technol. Adv. Mater. **20**, 1090 (2019).
[43] B. Ziaja, R. A. London, and J. Hajdu, Ionization by impact electrons in solids: electron mean free path fitted over a wide energy range, J. Appl. Phys. **99**, 033514 (2006).
[44] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Machine learning of molecular electronic properties in chemical compound space. New J. Phys. **15**, 095003 (2013).
[45] T. Ueno, H. Hino, A. Hashimoto, Y. Takeichi, M. Sawada, and K. Ono, Adaptive design of an x-ray magnetic circular dichroism spectroscopy experiment with Gaussian process modelling, npj Comput. Mater. **4**, 4 (2018).
[46] Y. K. Wakabayashi, T. Otsuka, Y. Taniyasu, H. Yamamoto, and H. Sawada, Improved adaptive sampling method utilizing Gaussian process regression for prediction of spectral peak structures, Appl. Phys. Express **11**, 112401 (2018).

[47] 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 É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

[48] H. Shinotsuka, B. Da, S. Tanuma, H. Yoshikawa, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths. XI. Data for liquid water for energies from 50 eV to 30 keV, Surf. Interface Anal. **49**, 238 (2017).

[49] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. (Morgan Kaufmann Publishers, San Francisco, 2016).

[50] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, London, 2006).

[51] C. Cortes and V. Vapnik, Support-vector networks, Machine Learning, **20**, 273 (1995).

[52] Z. H. Zhou, *Machine Learning* (Tsinghua University Press, Beijing, 2016).

[53] S. Tanuma, C. J. Powell, and D. R. Penn, Use of sum rules on the energy-loss function for the evaluation of experimental optical data, J. Electron Spectrosc. Relat. Phenom. **62**, 95 (1993).

[54] S. Adachi, *The Handbook on Optical Constants of Metals: In Tables and Figures* (World Scientific, Singapore, 2012).

[55] C. Wehenkel and B. Gauthé, Electron energy loss spectra and optical constants for the first transition series from 2 to 120 eV, Phys. Stat. Sol. **64**, 515 (1974).

[56] B. L. Henke, E. M. Gullikson, and J. C. Davis, X-ray interactions: photoabsorption, scattering, transmission, and reflection at $E = 50 - 30,000\,\mathrm{eV}$, $Z = 1 - 92$, At. Data Nucl. Data Tables **54**, 181 (1993).

[57] P. Prieto, F. Yubero, E. Elizalde, and J. M. Sanz, Dielectric properties of Zr, ZrN, Zr_3N_4, and ZrO_2 determined by quantitative analysis of electron energy loss spectra, J. Vac. Sci. Technol. **14**, 3181 (1996).

[58] D. E. Cullen, J. H. Hubbell, and L. Kissel, *EPDL97: The Evaluated Data Library*, 1997 version (Lawrence Livermore National Lab, Livermore, 1997).

[59] S.Tanuma, C. J. Powell and D. R. Penn, Calculations of electron inelastic mean free paths (IMFPs). VI. Analysis of the Gries inelastic scattering model and predictive IMFP equation, Surf. Interface. Anal. **25**, 25 (1997).

[60] F. Gerken, A. S. Flodström, J. Barth, L. I. Johansson, and C. Kunz, Surface core level shifts of the lanthanide metals Ce58-Lu71: a comprehensive experimental study, Phys. Scr. **32**, 43 (1985).

[61] F. Salvat, A. Jablonski, and C. J. Powell, ELSEPA—Dirac partial-wave calculation of elastic scattering of electrons and positrons by atoms, positive ions and molecules, Comput. Phys. Commun. **165**, 157 (2005).

[62] F. Offi and L. Petaccia, The attenuation length of low energy electrons in Yb, J. Phys.: Condens. Matter **22**, 305002 (2010).

[63] A. Jablonski, Universal quantification of elastic scattering effects in AES and XPS, Surf. Sci. **364**, 380 (1996).

[64] A. Jablonski and C. J. Powell, Effective attenuation lengths for photoelectrons emitted by high-energy laboratory x-ray sources, J. Elect. Spectros. Relat. Phenom. **199**, 27 (2015).

[65] S. Tanuma, C. J. Powell, and D. R. Penn, Calculations of electron inelastic mean free paths (IMFPS). IV. Evaluation of calculated IMFPs and of the predictive IMFP formula TPP-2 for electron energies between 50 and 2000 eV, Surf. Interface. Anal. **20**, 77 (1993).

[66] L. Bellaiche and D. Vanderbilt, Virtual crystal approximation revisited: application to dielectric and piezoelectric properties of perovskites, Phys. Rev. B **61**, 7877 (2000).