9P-13

Investigation of feature combinations for machine learning of projectile ion range

Hideaki Minagawa[1], Tomoya Tezuka[2], and Hidetsugu Tsuchida[2,3]

[1]Ion Technology Center Co., Ltd. 2-8-1, Tsudayamate, Hirakata, Osaka 573-0128, Japan
[2]Department of Nuclear Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8530, Japan
[3]Quantum Science and Engineering Center, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

Introduction

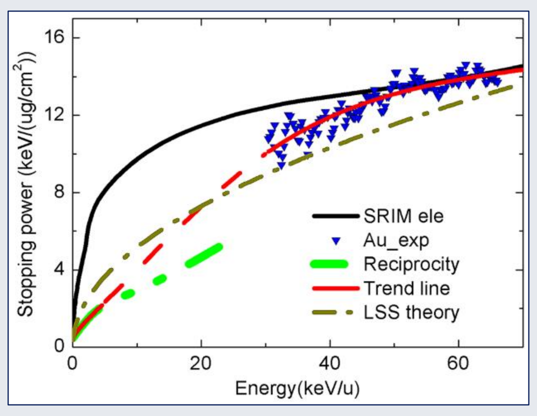
The process of ion implantation is crucial in producing semiconductor devices. To promote researches and developments or minimize prototyping costs, accurate simulations of implanted ion distribution is necessary.

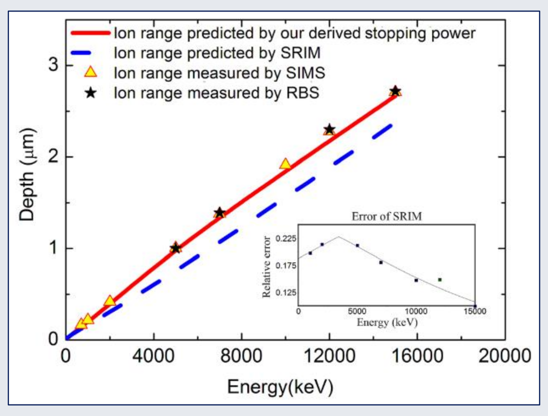
One widely used Monte Carlo simulation, such as SRIM, utilizes stopping power data from various experiments to calculate the distribution of implanted ions through processes such as ion scattering and energy loss.

However, there is a problem with the implanted ion range of the compound targets due to insufficient stopping power data. This issue leads to inconsistent experimental results in certain situations.

Inconsistent between SRIM prediction and experiments

SRIM prediction of ion implantation range R_p is smaller than experiments due to underestimation of the stopping power.





[1] K. Jin, et.al., J. Appl. Phys., 115 (2014) 044903

Purpose

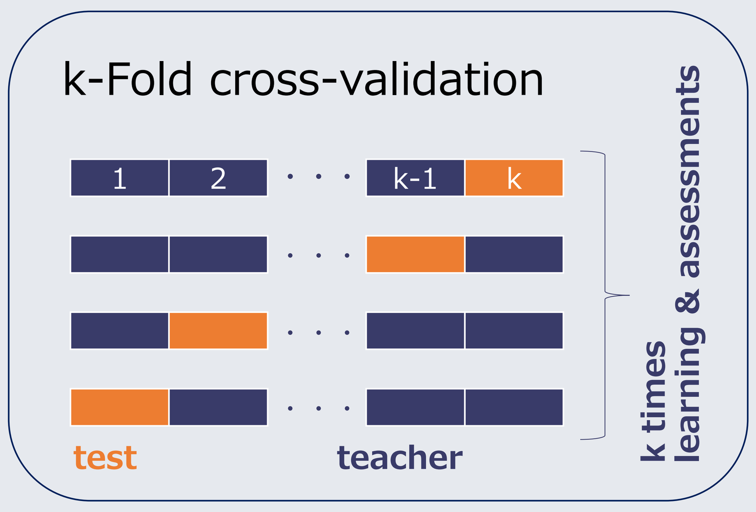
In this work, we used machine learning to create a model to predict the range of implanted ions R_p, **even when the stopping power is unknown**.

We also identified the key features necessary for accurate range prediction.

Methods

Data

- Compounds (A_xB_{1-x}: SiC, Al_2O_3, etc.)
  - 27 works (359 exp. data)
- Elements – SRIM
  - Solid target (uniform distribution of $Z_{1,2}$)
  - H – U ion



Feature sets for R_p predictions

Projectile

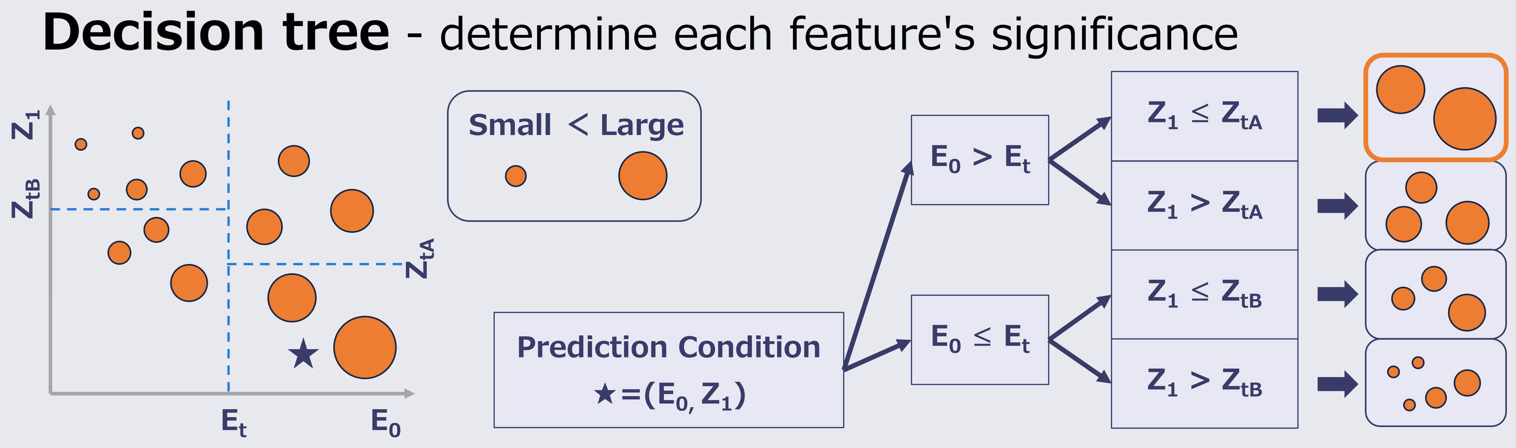
0: Incident energy E_0
1: Atomic number Z_1
2: Mass number m_1

Target

3: Atomic number $Z_2 = Z_{2A} \cdot x + Z_{2B} \cdot (1-x)$
4: Mass number $m_2 = m_{2A} \cdot x + m_{2B} \cdot (1-x)$
5: Density ρ
6: Mass difference $m_d = |m_{2A} - m_{2B}|$

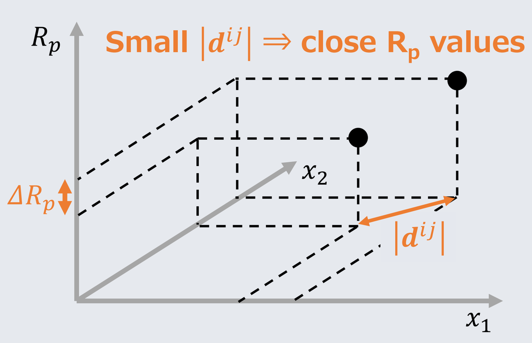
Machine Learning Model

- **Decision tree** - determine each feature's significance



The prediction value of R_p is the average value for each group

- **KRR** (Kernel Ridge Regression) with a Gaussian kernel
  - measure the correlation between the seven features and R_p



$$R_p = \sum_{i=1}^{N} w_i \exp\left(-\gamma \left|\boldsymbol{d}^{ij}\right|^2\right) : \text{Gaussian Kernel}$$

$$\left|\boldsymbol{d}^{ij}\right|^2 = \Delta x_1^2 + \Delta x_2^2 + \cdots + \Delta x_n^2$$

$$(x_i: E_0, Z_1, \ldots, m_d)$$

- Extensible to multi-dimension
- to prevent overfitting ⇒ ridge regression

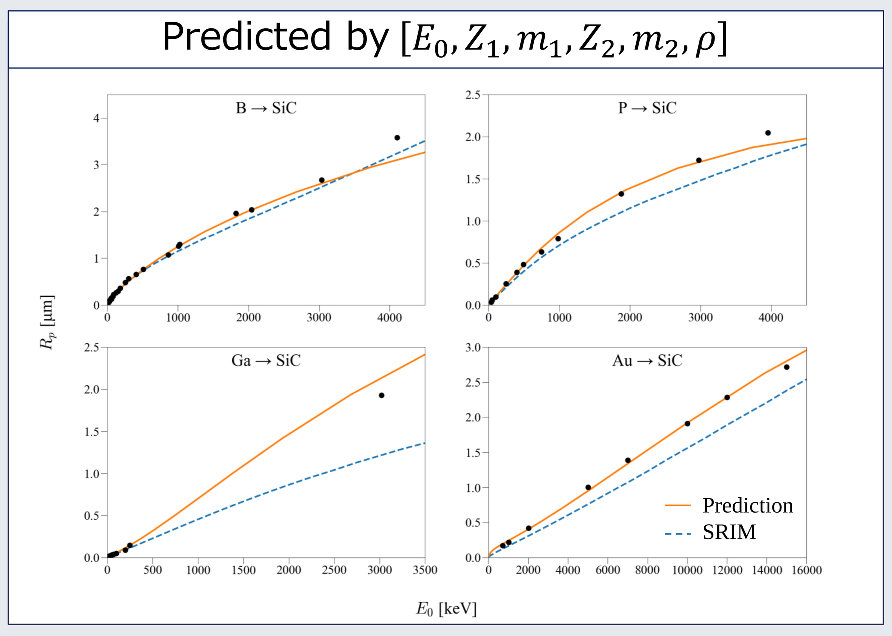
Assessment

the importance of feature combinations for the KRR model was examined by the coefficient of determination r_2.

$$r_2 = 1 - \frac{\sum_{test}\left(R_p^{test} - R_p^{pred}\right)^2}{\sum_{test}\left(R_p^{test} - \overline{R_p^{test}}\right)^2}$$

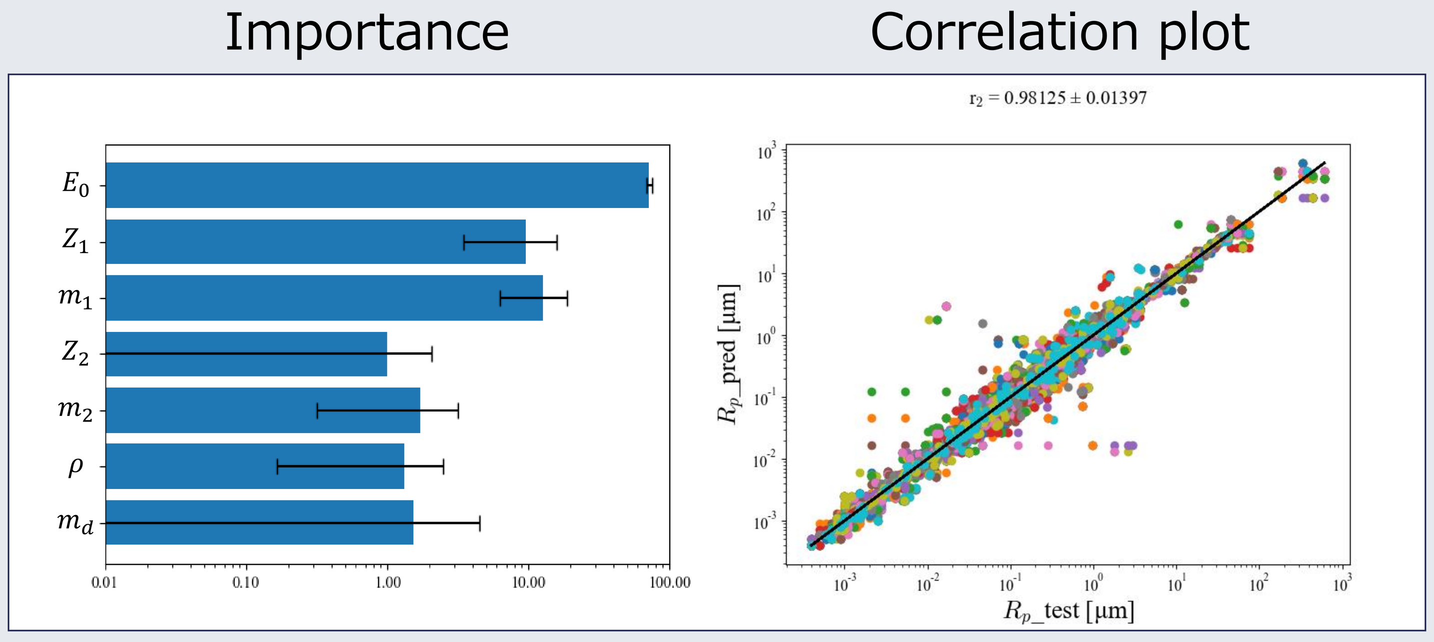
Summary

- Features related to projectiles and targets are indispensable for accurate prediction.
- This model can predict R_p as accurately as SRIM for any compounds without the stopping power data.



Results & Discussion

Decision Tree



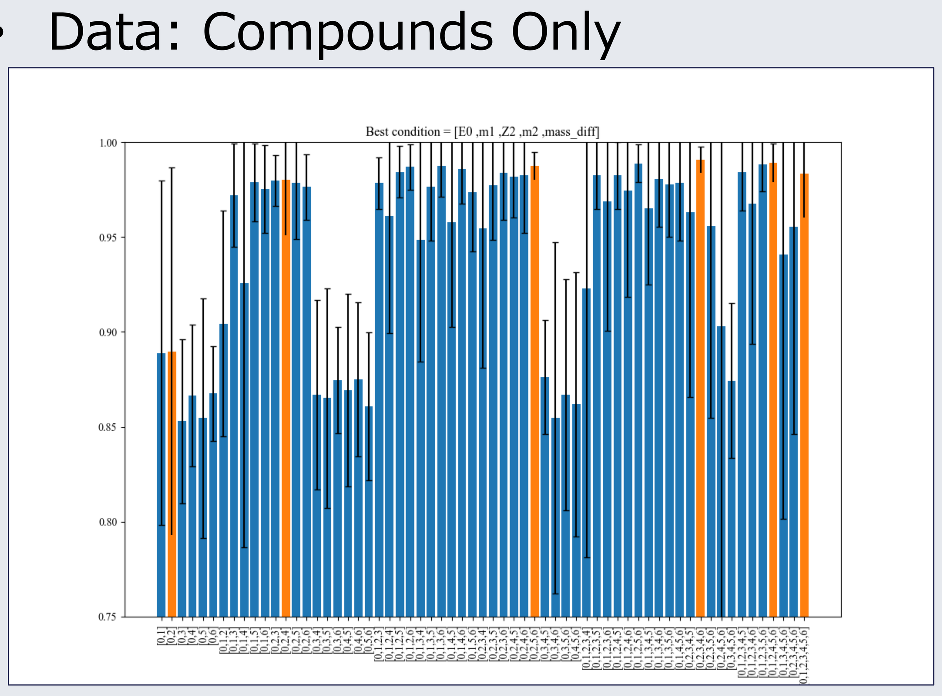
This model can determine the importance of each features, but the importance of the combination is unknown.

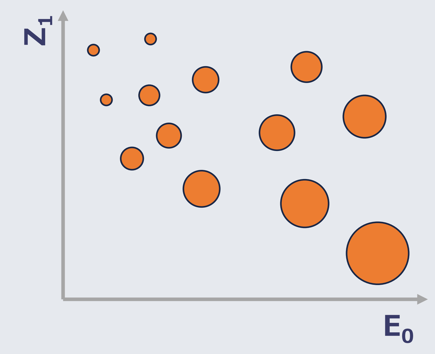
KRR model

$\overset{Z_1, m_1}{[1,2]}$ and $\overset{Z_2, m_2, \rho, m_d}{[3,4,5,6]}$ is necessary

Accuracy increase with dimensions.

- Data: Compounds Only

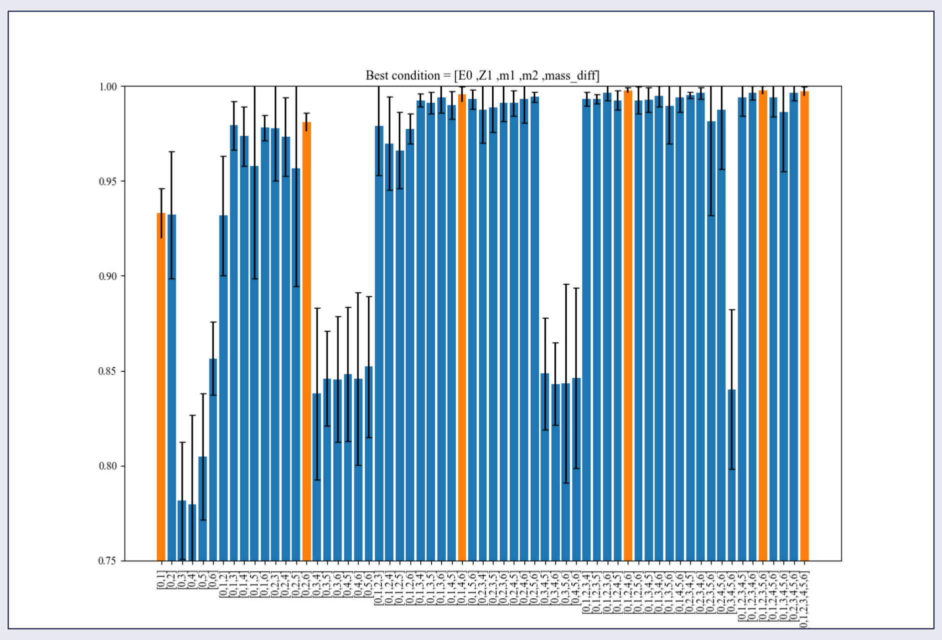


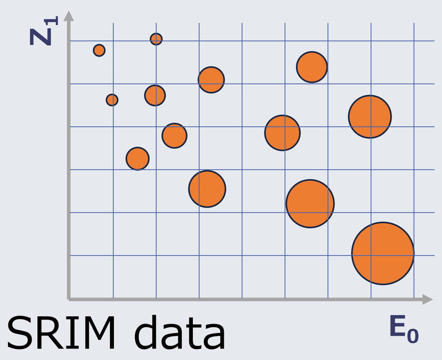


- The least accuracy is attributed to the small number of data points.

Intersection: SRIM data

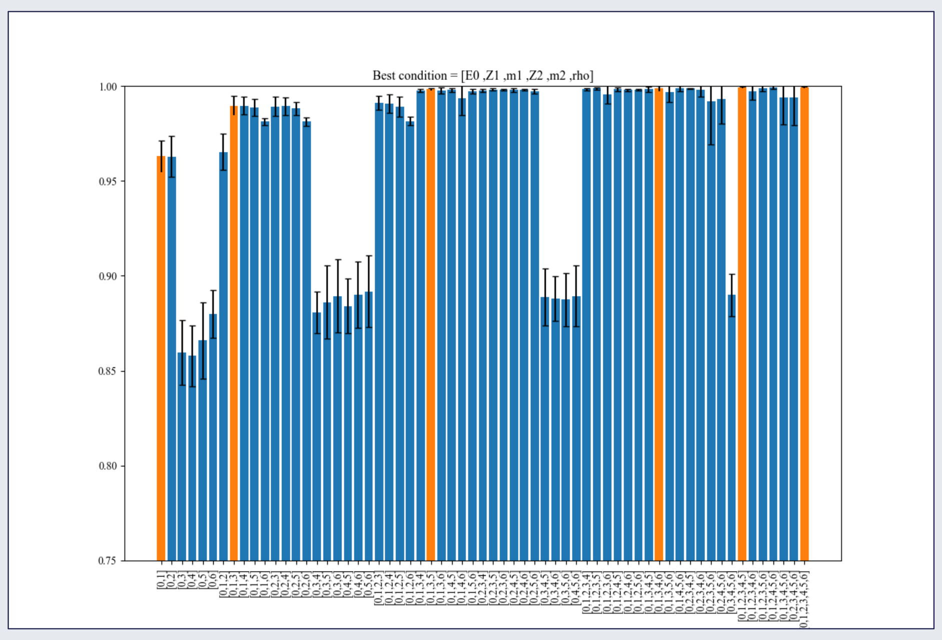
- Data: Compounds + Elements (Sparse)

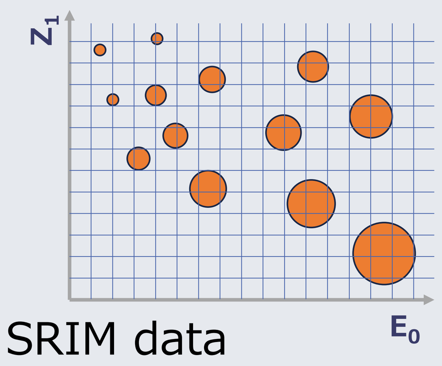




- SRIM data
  - 6 projectiles
  - 6 targets
- The accuracy increases by using SRIM data.

- Data: Compounds + Elements (Dense)





- SRIM data
  - 12 projectiles
  - 11 targets
- 6: Mass difference m_d shows relatively small contribution.
  - This is due to the highest SRIM data ratio.