“A Billion-Dollar Material”: Predicting Emittance of Photocathode Materials

Stanford | ENGINEERING
Computer Science

Michael Cai – CS ‘20
mcai88@stanford.edu

Michael Cooper – CS ‘20
coopermj@stanford.edu

Jennie Yang – MSE ‘19
jenniey@stanford.edu

STANFORD MATERIALS SCIENCE AND ENGINEERING

Motivation

- Lab of **Dr. Evan Reed** is collaborating with **SLAC** to screen **low thermal emittance photocathode materials** for particle accelerator applications.
- Current method based on **Density Functional Theory (DFT)** – computationally expensive.
- Our project aims to predict **high (0)** or **low (1) emittance** for each material.

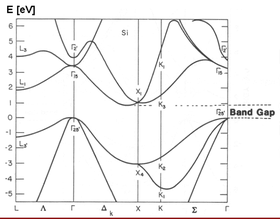

Figure 1: A sample electronic band structure, the output of DFT calculations and a crucial component to calculation of emittance. DFT is the primary computational bottleneck in the process of computing emittance.

Models

- **Logistic Regression:** $\mathcal{L}_2$ norm, 10^7 iterations maximum
- **SVM:** penalty of 1.0, RBF kernel, kernel coefficient $\frac{1}{num\_features}$
- **Naive Bayes:** Bernoulli model, without Laplace smoothing
- **Random Forest:** 100 trees, Gini impurity criterion
- **K-Nearest Neighbors:** $k = 2$, neighbors weighted by inverse of distance
- **Neural Network:** 7 hidden-layer fully connected neural network with Adam optimizer, CE loss, and learning rate = 0.0008

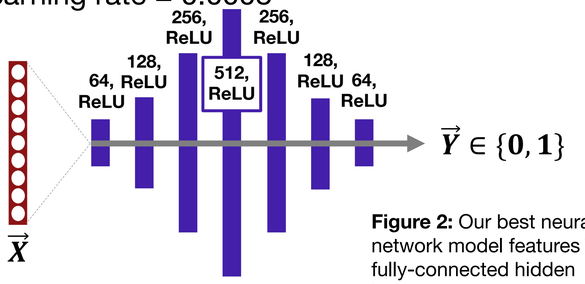

Figure 2: Our best neural network model features 7 fully-connected hidden layers

Data & Features

- The Reed Lab provided us with a labelled dataset consisting of **9,436 candidate materials** and their corresponding emittances (calculated via DFT). We shuffled and split the data **80/10/10** train/valid/test.
- Materials with minimum emittance ≤ **0.2** given the label **1** (low emittance, positive), else labelled **0** (high emittance, negative).
- Overall, we have **3276** positive and **6160** negative examples.

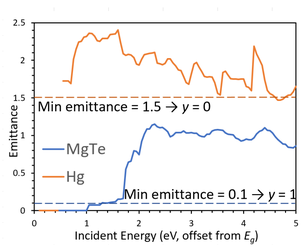

Figure 3 (left): Calculating Y labels for Magnesium Telluride (blue) and Mercury (orange)

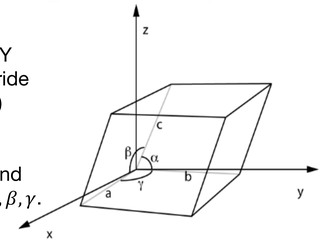

Figure 4 (right): A unit cell representation featuring bond lengths a, b, c and angles α, β, γ.

Unit Cell Representation:

$$\vec{X}_u = [a, b, c, \alpha, \beta, \gamma, v, e_0, a_0, b_0, c_0, \ldots, e_{15}, a_{15}, b_{15}, c_{15}]$$

where $a, b, c, \alpha, \beta, \gamma$ are lattice parameters (see Fig. 4), v is the unit cell volume, e_i is i^{th} atom’s atomic number, a_i, b_i, c_i are its coordinates

Average Properties Representation:

$\vec{X}_a$ = 16 elem. vector of avg. properties of atoms in material

Combined Representation:

$$\vec{X}_c = \text{concatenate}(\vec{X}_u, \vec{X}_a)$$

Results

Method	*val. acc*	*val. pr*	*val. re*	*val. f1*
LogReg	0.626	0.064	0.106	0.080
SVM	0.674	0.105	0.188	0.135
NB	0.552	0.230	0.336	0.273
RF	0.733	0.218	0.423	0.287
kNN	0.686	0.302	0.549	0.390
NN	0.668	0.331	0.585	0.423

Figure 5: Unit Cell Representation results by model

Method	*val. acc*	*val. pr*	*val. re*	*val. f1*
LogReg	0.639	0.081	0.137	0.102
SVM	0.685	0.130	0.235	0.167
NB	0.622	0.0	0.0	N/A
RF	0.749	0.273	0.541	0.363
kNN	0.676	0.268	0.479	0.344
NN	**0.748**	**0.333**	**0.658**	**0.442**

Figure 6: Average Properties Representation results by model

Method	*val. acc*	*val. pr*	*val. re*	*val. f1*
LogReg	0.667	0.162	0.286	0.207
SVM	0.642	0.036	0.062	0.046
NB	0.551	0.231	0.336	0.274
RF	0.758	0.237	0.476	0.317
kNN	0.708	0.302	0.566	0.394
NN	0.734	0.330	0.641	0.436

Figure 7: Combined Representation results by model

- Best results on the validation set obtained via **Average Properties** representation with **Neural Network** model.
- Test set *acc*: **0.701**, *pr*: **0.296**, *re*: **0.573**, *f1*: **0. 390**

Discussion & Future Work

Our model performs relatively well, but there is room for improvement…

- Our work shows that it is **possible** to **classify emittances** with a good degree of accuracy (over **70%**!) using a ML model.
- **Faster Classification** is possible using ML, taking minutes instead of hours – models can be used to prioritize which candidates to analyze.
- **Extremely complicated problem:** More powerful, physics-based models clearly perform better on this task, showing the difficulty of modeling quantum-mechanics problems with ML.
- **Insufficient data:** 9,436 data points may have been insufficient to fully understand the impact of each feature on the output.
- We need more **advanced data representations** for the structures and properties of the materials – this is area of active research.

Possible Future Extensions:

- Utilize **data augmentation** to increase the size of our dataset.
- Train a **generative model** to output the emittance graph or electronic band structure, rather than just a simple 1/0 classification of minimum emittance.

References

[1] D. H. Dowell et al, Cathode R&D for Future Light Sources, Nu-clear Instruments and Methods in Physics Research Section A, vol. 622, no. 3, pp. 685697, Oct. 2010.
[2] I. Bazarov et al, ”Thermal emittance measurements of a cesium potassium antimonide photocathode,” Applied Physics Letters, vol. 98, 02 Jun. 2011.
[3] T. Li et al, ”Emission properties of body-centered cubic elemental metal photocathodes,” Journal of Applied Physics, vol. 117, 10 Mar. 2015.
[4] L. Himanen et al, ”DScribe: Library of Descriptors for Machine Learning in Materials Science,” arXiv:1904.08875 [cond-mat.mtrl-sci] 18, Apr. 2019.
[5] E. Cubuk, et al “Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data” Journal of Chemical Physics, vol 150, 3 Jun. 2019.
[6] L. M. Mentel, Mendeleev - A Python resource for properties of chemical elements, ions and isotopes. , 2014