Supplemental Information

Mapping Pair Distribution Functions of Zirconium in NaF-ZrF4 Molten Salt from X-ray Absorption Spectroscopy Data via Machine Learning

Omar Oraby[1*], Nichollas Marcella[2,3], Anubhav Wadehra[4], Uday Pal[5], Karl Ludwig[6], Stephen Lam[1*]

[1] Department of Chemical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854, United States

[2]Department of Geosciences, Stony Brook University, Stony Brook, NY 11794, United States

[3]Joint Photon Sciences Institute, Stony Brook University, Stony Brook, NY 11794, United States

[4]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States

[5]Division of Materials Science and Engineering Boston University, Boston, Massachusetts 02215, United States; Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02215, United States

[6]Division of Materials Science and Engineering, Boston University, Boston, Massachusetts 02215, United States; Department of Physics, Boston University, Boston, Massachusetts 02215, United States

1. Conventional Fitting of the EXAFS Equation

While the fit presented in figure S1 appears reasonable, the extracted structural parameters summarized in table S1 indicate the presence of only a single coordinating atom in the first shell. This outcome was attributed to the fact that the EXAFS theoretical calculations that are used to fit the experiment predominantly capture the most significant single-scattering contributions from the nearest neighbor. In highly disordered environments such as molten salts, atoms in the first coordination shell are distributed over a range of distances. Consequently, the atom closest to the absorber dominates the EXAFS signal, leading to an apparent underestimation of the coordination number.

Table S1: Results of traditional EXAFS fitting for NaF-ZrF4 (57-43) %mol

N	S_0^2	σ^2	R
1.005 +/- 0.55	**0.971**	**0.0011100**	**1.88438**

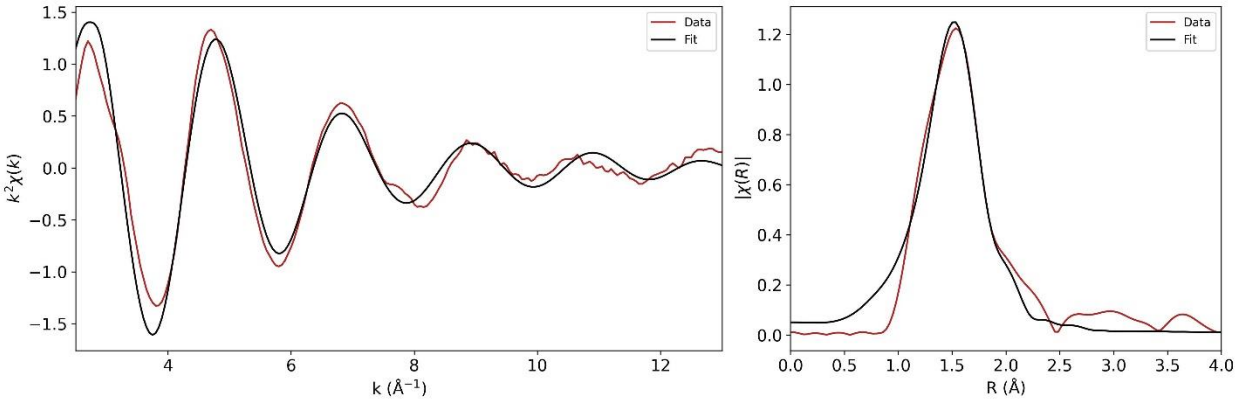

Figure S1: Conventional EXAFS fit in k-space (left pane) and its corresponding Fourier transform (right pane)

2. FEFF Calculations

Zr K-edge EXAFS spectra were calculated for the ONNE generated structures using FEFF10 code, where the amplitude reduction factor S_0^2 was set to 1.0, EXAFS k_{max} was set to 14 Å^{-1}, COREHOLE was set to Final State Rule (FSR) and RPATH was set to 6.0 Å. A self-consistent calculation with a 4 Å cluster radius was used. A bash script was generated to run multiple FEFF calculations simultaneously on a high-performance computer (HPC) cluster,

3. Convolutional Neural Network (CNN) Architecture

The CNN architecture consists of two sequential convolutional blocks. The first block takes input spectra of shape (1, 220) and applies 64 convolutional filters with a kernel size of 3, followed by batch normalization to stabilize learning and accelerate convergence. The Scaled Exponential Linear Unit (SELU) activation function was used to promote self-normalizing behavior and improve robustness of the CNN. This is followed by a max-pooling operation with a kernel size

of 2, which reduces the dimensionality and extracts dominant features while maintaining translation invariance. A dropout layer with a rate of 0.2 is included to prevent overfitting. The second convolutional block follows a similar pattern, expanding the number of filters to 128 while retaining the same normalization, activation function, pooling, and dropout layers. The output from the convolutional layers is then flattened and passed through a fully connected dense block, starting with a hidden layer of 768 units (3 × 256), again followed by batch normalization, SELU activation, and dropout. The final output layer is a linear layer sized to match the target number of structural outputs. The fine-tuned version was developed by appending additional dense layers to the base CNN. The output of the frozen base model, a 130-dimensional latent feature vector is passed through two fully connected layers of 256 units each, using SELU activation and dropout, followed by a final linear layer mapping back to 300 outputs. The use of SELU activation throughout the network ensures that the neuron activations remain approximately zero-centered and unit-normalized, reducing the need for explicit normalization in later layers and improving training stability.

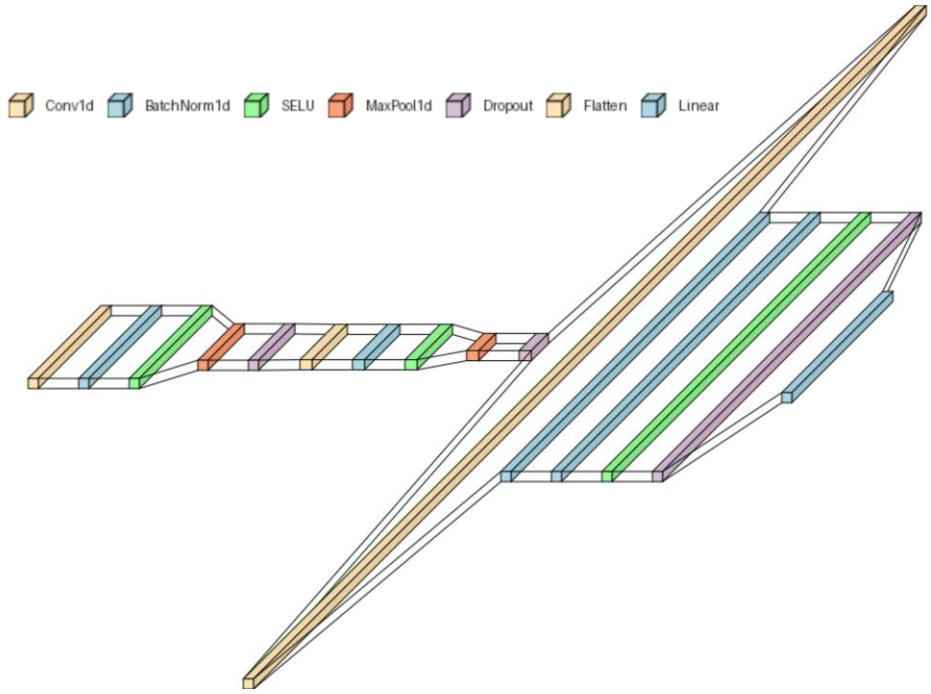

Figure 2: Base model CNN architecture used in this study

4. Training and Optimization Details

The model was trained using a batch size of 64 and optimized over 50 epochs using the Adam optimizer with a learning rate of 0.0001. A weight decay factor of 1×10^{-5} was used to introduce L2 regularization and penalize overly complex models. These hyperparameters were chosen based on empirical tuning across a validation set and informed by prior studies in similar regression tasks involving EXAFS inversion. A lower learning rate was selected to ensure stable convergence,

especially given the SELU activation function and the depth of the network. The batch size was set to balance between gradient estimate stability and training efficiency, while the number of epochs was chosen to allow the model sufficient time to learn without overfitting, as verified by early stopping criteria on validation error. The combination of dropout, weight decay, and batch normalization proved effective in mitigating overfitting and enhancing generalization, especially in the context of experimental noise and structural diversity inherent in molten salt systems.