$f$-electron structure database (FESD): A data analysis and learning tool for strongly correlated materials

Hasnain Hafiz, Adnan Ibne Khair, Hongchul Choi, Abdullah Mueen, Arun Bansil, Stephen Eidenbenz, John Wills, Jian-Xin Zhu, Alexander V Balatsky, and Towfiq Ahmed

1Department of Physics, Northeastern University, Boston, MA
2Department of Computer Science, University of New Mexico, Albuquerque, NM
3Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM
4Information Science and Technology Institute, Los Alamos National Laboratory, Los Alamos, NM
5Institute for Materials Science, Los Alamos National Laboratory, Los Alamos, NM

(Dated: October 5, 2017)

We developed a dedicated database and data driven computational tools for structural and electronic properties of $f$-electron based materials (e.g., lanthanide and actinide compounds). Due to a complex interplay among the hybridization of $f$-electrons to non-interacting conduction electrons, spin-orbit coupling, and local dynamical electron-electron interactions, no single model or first-principles based method can fully explain all the structural and functional properties in such systems. Complexity and richness in the electronic properties of $f$-electron compounds make them superior candidates for various technology and security applications. This calls out for the development of dependable predictive tools to aid the experimental search of new functionally correlated materials. Keeping this in mind, we adopted a data-driven approach. Using state-of-the-art algorithms and query tools, we train learning models from a large set of simulated data for the existing crystals of actinide and lanthanide. The machine learned models can then be used to search for new classes of systems with desired electronic and other physical properties. This can also be extended to search for new functional materials with desired stability. In the first phase of development, our database contains a large amount of simulated data and initial query tools. In this manuscript, we discuss the basic structure of the database. We also present some results for cleaning and correcting the structure data files. Electronic strong-correlated features from the high throughput density functional theory (DFT) simulation are also discussed.

I. INTRODUCTION

The world currently faces an ever increasing need for materials with newer electronic functionalities and higher energy efficiency. By integrating computational and experimental techniques, the materials science community has been desperately seeking for dependable predictive methods to gain insights into the structure-property relationships of functional materials. If achieved, such tools will not only make the materials research cost-effective, but will also accelerate searching for new materials with novel functionalities. Towards this goal, the search space for novel materials has also been augmented with compounds containing elements from the last rows of periodic table, e.g., lanthanides and actinides.

Although widely known for nuclear energy applications, actinide and lanthanide compounds have recently attracted attention for their complex and interesting electronic properties such as superconductivity, magnetism, and Kondo physics with numerous energy related applications. To leverage on these rich electronic properties, and to design new materials, one must have some predictive capability for enhancing structural, thermal and electronic functionalities which are often dictated by the electronic structure. While a 1-to-1 approach (one system - one calculation - one experiment) has been the traditional research strategy in lanthanide and actinide sciences, it is quite inefficient due to the high cost involved in handling high Z materials in experiments.

Theoretically, the complex electronic interactions from the localized $f$-orbitals also make it difficult to develop predictive insights for these compounds. This leaves the experimentalists to continue their search for new materials based on serendipity. To elaborate on these points, we note that the physics of $f$-electron systems is essentially a quantum many-body problem, and is notoriously hard to model, even with the state-of-the-art simulations techniques, e.g. density functional theory (DFT) methods. Even the best available dynamical mean field theory (DMFT) methods require, as an input, a semi-empirical correlation parameter $U$, which significantly reduces its predictive power from first principles. To address both these challenges, we developed a unique predictive model by adopting an integrated data and theory driven approach. We built a database with experimental and DFT simulated data. Using state-of-the-art algorithms, we trained our machine learning models, and also analyzed the simulated and experimental data. This helped us to reveal some of the most intricate electronic and structural properties emerging from the strongly-correlated $f$-electrons, and their hybridization with the rest of the itinerant electrons in the system.

In this paper, we briefly describe the structure and the basic work-flow of our database. As the first operational task, we collected all the existing crystal structures of $f$-electron compounds available in the Inorganic Crystal Structure Database (ICSD) and Crystallographic Open Database (COD), and organized them in our
database. The data-cleaning is a natural second step to correct errors in the collected data and to prepare them for theoretical simulations. This step involves machine learned predictions which are described in detail in Section IIIA. The cleaned data allows theoretical simulations with confirmed lattice parameters. This also provides the research community with complete crystallographic information files (CIF). Based on cleaned crystallographic data, we simulate a large amount of ground state electronic structure information using DFT. In this work, we present first-principles results for some of the most important f-orbital systems which contain Ce, Th, U, and Pu elements. We investigate the strongly correlated electronic properties from hybridization functions and orbital-projected band-structure calculations. We use DFT, DMFT and structured query language (SQL) tools to analyze the f-electron compound properties. We identify a list of compounds with a given range of f-orbital contribution in the band-structure. Moreover, we identify an anti-correlation between hybridization and volume which is in good accord with our earlier work for Ce based compounds\textsuperscript{14}. The details are presented in Section IIID.

One of the most unique features of our database\textsuperscript{8} is the high quality of simulated electronic structure data using DFT+DMFT\textsuperscript{7,15}. This method captures the strong-correlation and spin-orbit coupling in an all-electron environment. Many popular material databases (e.g. AFLOW\textsuperscript{16}, the Materials Project\textsuperscript{17}, the Organic Materials Database\textsuperscript{18}), use pseudo-potential and plane wave based techniques for DFT simulation. In contrast we used full potential LAPW method\textsuperscript{19–21}. Although computationally more intensive, it is very important to use all-electron relativistic calculations to yield a faithful predictive model for strongly correlated systems.

## II. DATABASE

Our database is primarily an electronic-structure database with query tools that allows to search for compounds with desired crystal and/or band structure information. In addition, the database contains DFT simulated data with multiple query capabilities.

Our database management system (DMS) is built on MySql and Java Development platforms. Additionally, several python based scripts are used for data parsing at various levels (e.g., CIF, DFT output). A few independent query tools are currently implemented in our database which are explained in Section III. As a work-in-progress, we are currently designing an application programming interface (API), which will integrate various query tools and provide a python based library of functions for programmed access to the database.

We discuss different components of our database in the following sections.

### A. Collection of Data

We acknowledge the ICSD and COD as our initial source of CIF files. Currently ICSD and COD contain approximately 188,000 and 364,331 CIF files, respectively. There are 54,465 structure files in ICSD and 27,502 structure files in COD with compounds containing lanthanide and actinide elements (e.g. outer shell f-orbital electrons). We parsed various structure information, such as crystal structure, lattice parameters, atomic positions etc. from these CIF files in our database.

While re-structuring the data format, we also developed multiple query tools with new search features which are not available in other existing databases\textsuperscript{16–18} to the best of our knowledge. For example, we provide extended search capabilities based on various new functionalities, such as space group or lattice system based search, and superstructure search with desired symmetry.

As part of this development process, we also performed the ‘data cleaning’ step by employing different supervised learning tools. These resolved the ambiguities in the CIF files due to the missing information, and corrected the structure files to be used as input data for the DFT simulation. Details of the data cleaning and verification steps are described in Section IIIA.

The cleaned and verified structure information gives us the platform for generating reliable DFT simulated electronic structure data (e.g. band structure, DOS). This is the core content of our f-electron database. Our data generation also involves \textit{ab initio} calculations in a high throughput manner on a parallel computing platform. Details of these calculations are discussed in Section IIIB.

### B. Generation of Data

The electronic ground states are obtained using density functional theory (DFT) with full potential linearized augmented plane-wave (FP-LAPW) scheme as implemented in WIEN2k package\textsuperscript{21}. To account for the exchange and correlation effects, we used generalized gradient approximation (GGA) with Perdew-Burke-Ernzerhof (PBE) functional\textsuperscript{22}. The calculation started with the transformation of CIF files into WIEN2k input files, using the tool ‘cif2struct’. During this preparation stage, the space group symmetry of each system was correctly captured based on all atomic positions. We performed self consistent calculations, using a sufficiently large plane wave cut-off (\( \text{RKmax} = 9.0 \)), and a \( 10 \times 10 \times 10 \ k\)-mesh for the first Brillouin zone. The ground state energy and electronic charges were well converged in our calculations. For band-structure calculation, we considered the standard high-symmetry \( k\)-path of respective crystal and space group symmetries. The density of states (DOS) was obtained over a uniform energy mesh and resolved into the basis of real orbitals. The hybridization functions \( \Delta(E) \) for \( 4f \) and \( 5f \) electrons were calculated solving the local Green’s function
on the real axis where a single shot dynamical mean-field theory (DMFT) calculation was performed with the DFT+DMFT approach as implemented in Ref.\textsuperscript{7}. Since we are only interested in total $\Delta(E)$, a real harmonics basis with spin and without any symmetry was considered for the projection of the local Green’s function. For the accurate estimation of $\Delta(E)$, an energy window was set to $-20$ eV to $+10$ eV around Fermi level with a sufficiently large $k$-mesh, and a Lorentzian broadening of 0.25 eV. To correctly capture the trends over large data sets, all calculations were maintained at the same level of uncertainties.

III. APPLICATIONS OF DATABASE:

A. Data cleaning and verification

In our SQL database, some crystallographic features (e.g. lattice parameters, space group) are missing or ambiguous in thousands of compounds because of incomplete original CIF files that were parsed to create the database. Such systems cause convergence problems or inaccurate results in DFT simulations. To resolve this issue, we adopt a machine learning approach for verifying and cleaning the incorrect crystal information.

We investigate three supervised machine learning algorithms to predict the one of the seven possible ‘lattice systems’. Currently we have hosted crystal data for approximately 82,000 $f$-electron compounds, out of which 8,711 compounds contain missing or incorrect lattice systems. We use Logistic Regression (LR) that predicts categorical target variables using a logistic function on the linear combinations of feature values. K Nearest Neighbor (KNN) algorithm predicts category of an unknown instance based on K ‘similar’ instances. We use Euclidean distance as a measure of similarity and pick the majority of the three (K=3) most similar materials to predict the lattice parameters of an incomplete system. We also use Multilayer Perceptron (MLP) which is a forward feeding network of perceptrons, as opposed to a single perceptron that is equivalent to a logistic regression model. We use three layers of perceptrons, eight in each layer. Technical details of various learning methods, including parameter configurations and data format, are discussed in the supplementary materials.

To train these models, we use a set of 275,926 labeled instances from the COD database\textsuperscript{13} with accurate lattice constants $a,b,c,\alpha,\beta,\gamma$, volume, and space group as training features. We ran a 10-fold cross validation, and calculated the prediction accuracy, precision, and recall. Using the miss and hit between the predicted class and true class outcome, we constructed a representation widely known as the confusion (or error) matrix in the machine learning community. As shown in Fig.1, the confusion matrices for our three adopted methods clearly indicate the superior outcome from MLP, which essentially is a neural network approach. MLP has an accuracy of 99.1% in predicting the crystal system, as quantitatively shown in Table.1. Although computationally more complex and expensive, MLP clearly became our algorithm of choice, to clean and obtain high quality structure information for our database.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>90.4%</td>
<td>76.5%</td>
<td>85.6%</td>
</tr>
<tr>
<td>KNN</td>
<td>98.1%</td>
<td>87.6%</td>
<td>86.2%</td>
</tr>
<tr>
<td>MLP</td>
<td>99.1%</td>
<td>90.8%</td>
<td>91.3%</td>
</tr>
</tbody>
</table>

TABLE I. Benchmarking performance of learning techniques.

However, in all three models, we observe some off-diagonal non-zero values in the confusion matrix indicating errors in prediction. Interestingly, we find such incidents to be restricted to three unique classes, e.g., Trigonal, Hexagonal, and Rhombohedral (see Fig.1.a).
FIG. 2. Superstructure compatibility prediction of double perovskites $AA'BB'C_3C'_3$. Each data point is a paired combination of two single perovskite system as shown in the inset of (b). Different colors imply different combination of $C$ and $C'$ as shown in the legend of (b). Space group for each $C/C'$ type is given in the parenthesis of the legends of (b).

In crystallographic convention, there are two equivalent systems, e.g. crystal system and lattice system. Trigonal and Hexagonal are part of the Crystal system convention, the Rhombohedral and Hexagonal are defined in Lattice system. We observe such definitional ambiguities and mixing in different crystallographic systems in the original CIF files. Using our neural network (Fig.1(a)) based MLP tool, we are now able to resolve such ambiguities with overall higher accuracy, and thus, can provide a cleaner data set for further ab initio simulations.

Using the multilayer perceptron, we predicted the missing crystal systems for 8,711 compounds with the following distribution:

<table>
<thead>
<tr>
<th>Label</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>cubic</td>
<td>1355</td>
</tr>
<tr>
<td>hexagonal</td>
<td>4577</td>
</tr>
<tr>
<td>monoclinic</td>
<td>1428</td>
</tr>
<tr>
<td>orthorhombic</td>
<td>23</td>
</tr>
<tr>
<td>tetragonal</td>
<td>2</td>
</tr>
<tr>
<td>trigonal</td>
<td>1326</td>
</tr>
<tr>
<td>Total</td>
<td>8711</td>
</tr>
</tbody>
</table>

B. Superstructure compatibility prediction

Here, we demonstrate another application of our database after cleaning and confirming all the crystal structure files as described in Section IIIA. It is a common practice in material synthesis experiments to try to design new ordered superstructures with desired electronic properties by overlaying two or more compounds with different chemical composition but similar crystal symmetry. To provide guidance to such experiments, we were able to identify a set of perovskite compounds in our database, which possesses the maximum likelihood of forming paired and ordered superstructures. Although a careful and rigorous structural and electronic relaxation is an essential final step to confirm the existence of such structures, we realize that it is important and very efficient to first narrow down the search space before performing any costly simulation. Thus, using only the primary parameters (lattice parameters), and chemical intuition based on ‘ionic radius and tolerance’ arguments, we performed a data-mining query, and identified sets of paired superstructure combinations that have the highest likelihood of forming ordered superstructures.

Figure 2 shows a compatibility test for ordering of superstructures for a number of double perovskites with chemical formula $AA'BB'C_3C'_3$. Here, $A/A'$ site cations are either rare earths or actinides, $B/B'$ cations are transition metal elements, and $C/C'$ anions are mostly oxygen or halogens.

Using the ionic radii of two existing single perovskites $ABC_3$ and $A'B'C'_3$ (see inset of Fig. 2(b)), we predict the geometric stability of a possible double perovskite $AA'BB'C_3C'_3$ by calculating Goldschmidt tolerance factor:

\[ t = \frac{r_A + r_C}{\sqrt{2(r_B + r_C)}} \]

where $r_A$, $r_B$, and $r_C$ are the average ionic radii of $(A, A')$, $(B, B')$, and $(C, C')$ respectively, while neglecting other contributing factors such as octahedral tilting. Our superstructure compatibility test successfully predicted some of the existing double perovskites as shown in Table II (top). These compounds have high tolerance factors and can be synthesized at ambient pressure. In addition, we predicted some of the possible double perovskites that have high tolerance factors and need further experimental verification. A few of our prediction is given in Table II (bot-
FIG. 3. Web interface of the band structure of UN displaying f-orbital characters near Fermi level.

<table>
<thead>
<tr>
<th>Perovskite</th>
<th>Tolerance (t)</th>
<th>c1/c2</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>La2CoMnO6</td>
<td>0.96</td>
<td>0.98</td>
<td>26, 27</td>
</tr>
<tr>
<td>La2CrCoO6</td>
<td>0.97</td>
<td>1.02</td>
<td>28</td>
</tr>
<tr>
<td>La2CoFeO6</td>
<td>0.98</td>
<td>1.01</td>
<td>29, 30</td>
</tr>
<tr>
<td>La2GaMnO6</td>
<td>0.96</td>
<td>0.99</td>
<td>31, 32</td>
</tr>
<tr>
<td>Ce2AlCrO6</td>
<td>0.98</td>
<td>1.01</td>
<td>This work</td>
</tr>
<tr>
<td>Ce2GaCrO6</td>
<td>0.96</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>CeSmAlCrO6</td>
<td>0.96</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>LaCeAlMnO6</td>
<td>0.98</td>
<td>1.02</td>
<td></td>
</tr>
</tbody>
</table>

TABLE II. Predicted double perovskites from superstructure compatibility test, containing already synthesized (top) and newly predicted (bottom) compounds.

Considering each data point as a paired combination of double-perovskite superstructure, we plotted their tolerance factor $t$, lattice parameters ratio $\frac{a_1}{a_2}$ and $\frac{c_1}{c_2}$ in Figure 2(a), where $(a_1, c_1)$ and $(a_2, c_2)$ are lattice parameters of $ABC_3$ and $A'B'C'_3$ perovskites respectively. Our analysis shows clustering of data which stems from the correlations between the type of anion/cation and space group of the individual perovskites with tolerance and lattice parameters. We see a general trend of high stability ($t$) when both $C$ and $C'$ anions are oxygen or hydrogen, in contrast to the case when both of these anions are halogens. Our analysis also reveals that a high tolerance factor can be achieved when $\frac{a_1}{a_2}$ and $\frac{c_1}{c_2}$ are close to unity. In other words, when the difference of ionic radii of A ($= r_A - r_{A'}$) or B ($= r_B - r_{B'}$) site cations respectively is small. We also see a vital role of space groups such as two cubic perovskites usually results in a stable double perovskite in contrast to the cases when groups are mixed. This dependency of crystal structure comes naturally because similar crystal structures of two contributing perovskites produce less cation size mismatch or smaller difference in ionic radii of the cations of individual sites. Our analysis also finds a linear trend and grouping, when the double perovskite data points are plotted in the $\frac{a_1}{a_2}$ and $\frac{c_1}{c_2}$ plane as shown in Fig. 2(b). The central line indicates that most abundant double perovskite systems can be formed from $ABC_3$ and $A'B'C'_3$ cubic systems. Other linear trends show the cases when mixed systems can be formed from cubic and other types (non cubic) of perovskite crystal structures.

In recent years, there has been a lot of theoretical and experimental interest in creating layers and superstructures by-design, due to their enormous potential in achieving various emergent functionalities (e.g., superconductivity\textsuperscript{34,35}, multiferroics\textsuperscript{36,37}). Towards this effort, our $f$-electron database can provide the necessary ingredients including crystal and simulated data with relevant query tools, and thus, can significantly reduce the search space prior to any expensive computations.

C. Orbital resolved band structure search

DFT simulated electronic data (e.g., band-structure, DOS) make up the core content of our $f$-electron database. Band-structures, which essentially contain the energy-momentum dispersion relations of a system, often provide the most interesting signatures, which dic-
FIG. 4. The calculated hybridization function $\Delta(E)$ for a series of $4f$ and $5f$ monopnictides, $LnX$ and $AnX$ ($Ln = Ce, An = Th, U, Pu$ and $X = N, P, As, Sb, Bi$).

tate the microscopic properties of materials. Although very informative, interpretations of various features in the band-structures can be complex and nontrivial, particularly for strongly-correlated systems with prevalent many-body effects. We note that most compounds in our database are strongly-correlated due to the presence of localized $f$-orbital electrons at the valence level. The atomic and orbital momentum ($s$, $p$, $d$, and $f$) contributions along with the shape of the band-structures near the Fermi energy provide vital information for understanding the electronic behavior in these materials. Keeping this in mind, we have implemented a unique band structure search capability based on different atomic partial orbitals for each band, and at each k-point, along the high symmetry direction. As an example, in Fig.3 we show an advanced band structure query on Uranium Nitride (UN) that has dominant $f$-orbital contribution near Fermi energy. The bands near the Fermi level contain large $f$ characteristics, and their flatness indicates the atomic orbital-like localization.

To understand many complex electronic, magnetic, and optical properties of $f$-electron materials, such information can be extremely useful for developing machine learning models and predictive tools. Many such query tools are currently being developed and will be discussed in future work. However, we acknowledge that local dynamical correlation is missing in DFT simulated results in our database. Such limitations can be addressed with $ab\ initio$ dynamical mean field theory (DFT+DMFT), as discussed in the next section.

D. A search for strongly correlated actinides system

At the current stage of development, we have performed zeroth-order DMFT calculations on top of our ground state DFT calculations. This step captures frequency dependent hybridization functions for all the compounds in our database. In earlier work$^{14}$, we have found hybridization a good descriptor for detecting a localization trend. We acknowledge that an accurate description of strong-correlation and localization phenomenon can only be captured with self-consistently converged DMFT calculations. However, our findings$^{14}$ indicate that a high throughput zeroth-order DMFT based hybridization is quite capable of revealing the interesting qualitative trend. The maximum hybridization value near Fermi energy consistently decreases with increasing lattice spacing on the set of $\approx 300$ Ce based binary compounds.

Such a descriptor can provide useful insights for identifying the systems with dominant strong-correlation characters, and thus, can narrow down the search space. Using hybridization, crystal symmetry information, and band-structure based query, a smaller set of systems can be selected which will open up the pathway for a more computationally intense DMFT based investigation or
The most commonly used formalism for DMFT is based on the Anderson impurity model\textsuperscript{38} where an impurity electron is immersed in a sea of itinerant electrons - in our case the \( f \)-electrons which hybridize with the Bloch states of the surrounding electron sea\textsuperscript{39} to form a resonance described by

\[
H_{\text{resonance}} = \sum_{k,\sigma} \epsilon_k n_{k,\sigma} + \sum_{k,\sigma} [V(k) c_{k,\sigma}^\dagger f_{\sigma} + V(k)^* f_{\sigma}^\dagger c_{k,\sigma}]
\]

Here, the hybridization matrix \( V(k) = \langle k | V_{\text{atomic}} | f \rangle \) is basically the Bloch vector (\( k \)) averaged potential matrix elements between the \( f \)-orbitals and the itinerant electrons near the Fermi energy. This is defined as,

\[
\Delta(E) = \pi \sum_k |V(k)|^2 \delta(\epsilon_k - \mu) = \pi V^2 \rho(E)
\]

where \( V^2 \) is the average of the hybridization around the Fermi surface. A bigger overlap of \( f \)-states with the surrounding itinerant electrons will produce a larger hybridization function, and can serve as a descriptor for the localization/delocalization of \( f \) electrons.

In Fig. 4, we present the calculated hybridization function \( \Delta(E) \) for various \( 4f \) and \( 5f \) monopnictides \( LnX \) and \( AnX \) (\( Ln = Ce, An = Th, U, Pu \) and \( X = N, P, As, Sb, Bi \)) with space group \( Fm\overline{3}m \) and a structure similar to \( NaCl \). Our results for \( CeX \) clearly reproduce previously reported results of \( \Delta(E) \)\textsuperscript{14} and established trends in similar compounds\textsuperscript{40}. This also confirms the high fidelity nature of our calculation. Our results also identify many similarities in \( \Delta(E) \) features while comparing different \( 4f \) and \( 5f \) monopnictides. For example, a distinct peak in \( \Delta(E) \) is observed between 2 eV and 3.5 eV below the Fermi level for \( LnN \) and \( AnN \). This peak decreases monotonically within small fluctuations as we scan from the top to the bottom of group 15 of the periodic table. The reduction of \( \Delta(E) \) bandwidth for \( Sb \) and \( Bi \) compounds indicates a more localized nature of the \( f \)-electrons, while the sharp peak of the \( N \) compounds with a two times bigger area under the curve shows a higher degree of delocalization. Comparing \( 4f \) and \( 5f \) monopnictides, as shown in Fig. 5 (top), we see an overall increase of \( 5f \) hybridization function. In other words, because of the extended nature of the \( 5f \) shells, we see an decrease of localization for \( 5f \) monopnictides compare to their \( 4f \) counterparts. In addition, within \( 4f \) or \( 5f \) shell, if we move left to right of the periodic table, we see a increase in localization with increase in \( f \)-electrons. Moreover, we find an anti-correlation between the degree of localization and lattice constants for both \( LnX \) and \( AnX \) as shown in Fig. 5 (bottom). Moving from top to bottom of the group 15, we observe higher unit cell volume produces smaller \( \Delta(E) \) at Fermi level.

This can be understood from the fact that a larger ligand distance between \( f \)-electron atom and neighboring atoms decrease the amount of interactions between them, and thus, forms a more localized band structure. This picture is also true if one goes from \( 4f \) to \( 5f \) shells where the \( 4f \) monopnictides have overall higher volumes in contrast to their \( 5f \) counterparts, and show an increase in localization.

Based on hybridization analysis, one can equivalently look for other different features from DOS and bandstructure, in conjunction with the lattice symmetry information. Given the complexity of the systems of interest, we emphasize that a predictive model must narrow down the search space as a first step, by using domain knowledge from the hybridization features of a larger set of compounds. Only then, a reliable confirmation of the desired electronic properties using more sophisticated techniques (e.g. charge self-consistent DFT+DMFT calculation) will be practical, and help us to successfully achieve theory guided experimental discovery of new functional correlated materials.

IV. CONCLUSION

In this work, we presented our newly developed \( f \)-electron structure database (FESD). We also discussed and analyzed some selected preliminary results on correlated \( f \)-electron systems which are obtained using the database query tools.

At the moment, our database is limited to only existing and experimentally synthesized crystal structures. We further cleaned the data and confirmed their accu-
We observe an anti-correlation behavior between the hybridization functions between order DMFT on top of DFT calculations, we obtained hybridizations of strongly correlated materials. Using zeroth can be used to study the fundamental electronic interactions data (e.g., band structure, DOS) obtained from DFT simulation. We developed several query tools to analyze the atomic-orbital characteristics of the bands near the Fermi energy. Finally, we demonstrate that our database can be used to study the fundamental electronic interactions of strongly correlated materials. Using zeroth-order DMFT on top of DFT calculations, we obtained hybridization functions between \( f \) and itinerant electrons. We observe an anti-correlation behavior between the hybridization function and volume of the unit cell, which helps us identify the strongly-correlated systems from a wider range of compounds.

With all these different data query and mining tools, FESD database is uniquely attributed with high quality DFT simulated electronic structure information. Thus, our database is focused and well equipped for the discovery of next generation functional materials with strongly-correlated and localized \( f \)-electron character.

V. ACKNOWLEDGEMENT

This work is supported by the U.S.DOE at LANL under Project No. 20170680ER through the LANL LDRD program, and by the Institute for Materials Sciences (IMS), NSEC at LANL. We also thank Robert S. Markiewicz, Avadh Saxena, Heike Harper, Anil Kumar, and Prasanna Balachandran for useful discussion.

---

* hafiz.h@husky.neu.edu
† atowf@lanl.gov

24 V. M. Goldschmidt, Naturwissenschaften 14, 477 (1926).
(2005).