

Materials informatics as a computational infrastructure for materials discovery

P V Balachandran and K Rajan

Department of Materials Science and Engineering, Iowa State University, Ames, IA 50011

Email: krajan@iastate.edu

Abstract. One of the critical barriers in transitioning new materials development into engineering practice is the uncertainty associated with the data that is used in the materials design process. While constitutive modeling strategies still form the foundations for computational materials design, uncertainty and incomplete information still pervade. Often the available data space is limited, and the challenge of using that data to develop computational studies for exploring new materials is made difficult. In this paper we describe how the application of information entropy metrics coupled with the tracking of the statistics of data evolution can aid in identifying key structure-property relationships in materials. This information in turn can aid in identifying new materials chemistries with targeted properties narrowed down from a large and sparse chemical search space. A brief example in the discovery of new solid-state electrolytes for fuel cell examples is given as a template for this informatics-aided computational infrastructure.

1. Introduction

“One of the continuing scandals in the physical sciences is that it remains impossible to predict the structure of even the simplest crystalline solids from knowledge of their composition” [1]. This provocative observation—as well as similar ones made by other materials scientists over the past 50 years [2-5]—captures one of the great questions in materials crystal chemistry, namely, *why do atoms arrange themselves in the way they do?* The corollary to this question is the following: *Given that the underlying crystal structure defines many fundamental properties of materials; can we identify the chemistries and the associated crystal structure to define specific characteristics of materials so that we can use this information to create new materials with specifically tailored properties?* Despite advances in theory and computation, we still do not

have sufficiently robust physical models that can span enough chemistries to explain how atoms interact. To understand the complexity of the problem, consider the 76 useful stable elements in the periodic table. There are 2,850 binary, 70,300 ternary, 1,282,975 quaternary, and $> 10^9$ heptanary combinations of these 76 elements. It would be almost impossible to model all permutations of atomic-scale interactions, let alone experimentally synthesize different chemistries. This situation, in essence, defines the challenge that materials scientists face in systematically designing new materials chemistries, namely, exploring an unwieldy large chemical space with relatively little data that is known. We can, however, draw some guidance to address this issue by looking at systems biology, where the combination of systems-level experimentation and the use of quantitative and computational tools to integrate, visualize, and analyze the resulting experimental data has made it possible to systematically identify and characterize molecules and the molecular interactions that define cellular pathways, tissues, organs, and organisms. This allows us to obtain a quantitative and predictive understanding and derive solutions to fundamental biological problems based on dynamic relationships between genetic, molecular, cellular, physiological, and environmental factors.

Similarly, unraveling the complex relationships between structure, bonding, and chemistry in inorganic materials—for example, identifying pathways that demonstrate how parameters describing electronic structure, chemistry and crystal geometry “communicate” with each other to ultimately define properties—is needed to achieve significant advances in materials science. Through many seminal papers, Sir Alan McKay has expounded on the idea of a framework for “generalized crystallography” [6]. He has proposed that the crystal is a “structure the description of which is much smaller than the structure itself” and that this description of structure serves as a “carrier of information” about the structure on larger length scales [7,8]. He also suggested that these components of description of structure can help develop a “biological approach to inorganic systems” and proposed the construction of an “inorganic gene”. This serves as the motivation underlying the present study by exploring how fundamental pieces of information, treated as discrete bits of data, can collectively characterize the stability and properties of a given crystal chemistry. We show how the use of statistical learning tools, including fundamental concepts borrowed from information theory, can be used to characterize a crystal structure in terms of fundamental descriptors of information (i.e., the “genes”) and how these pieces of information are “sequenced” to guide the characteristics of that crystal structure and in fact help guide the development of new crystal chemistries.

The challenge in defining the gene in inorganic crystal chemistry is to characterize the appropriate combination of discrete characteristics associated with crystal chemistry that collectively define a particular property or set of properties of the material. Normally structure-property relationships are guided by defined functional relationships (e.g., electronic structure calculations to define energy landscapes associated with crystal chemistry). In this paper we describe an approach to establish such a structure-property relationship where we do not assume any specific formulation linking structure with property. Rather we take a data-driven approach where we seek to establish structure property relationships by identifying patterns of behavior between known discrete scalar descriptors associated with crystal and electronic structure and observed properties of the material. From this information we extract design rules that allow us to systematically identify critical structure-property relationships resulting in identifying in a quantitative fashion the exact role of specific combination of materials descriptors (i.e. genes) that govern a given property. This is the foundation of the concept of the quantitative structure-activity (or property) relationship, QSAR/QSPR, widely used in the field of organic chemistry

and drug discovery. We have shown the potential impact of applying this strategy for identifying new high-temperature piezoelectric materials [9].

In this paper, we demonstrate this approach by striving to track which combination of parameter appear to influence crystal stability by partitioning high-dimensional data sets. Using the concept of information entropy (e.g., Shannon entropy) as a selection criterion, we propose a high-dimensional recursive partitioning strategy to develop classification schemes that previously have been approached by empirical observations. Thus, the aim of the classification is to reduce the uncertainty of the dataset and to track precisely which and how many latent variables contribute to structural stability. Using such information content measures, we will be able to *quantitatively* rank the influence of a vast array of parameters.

2. Exploring the chemical search space

Increasingly, scientists need to discover new oxide-ion conducting materials, since such materials form the basis of a range of important, environmentally friendly applications such as sensors, gas separation membranes and solid oxide fuel cells (SOFCs). In particular, the impact of SOFCs has held the greatest potential for next generation power production [10]. Normally, SOFC systems are operated in the high temperature regime of 850-1000°C [11]. The operating temperature is influenced by the nature of application and properties of available solid electrolyte. One of the well-known solid electrolytes that is widely employed in SOFC applications is yttria-stabilized zirconia (YSZ), which when operated at temperatures greater than 800°C exhibits high oxide-ion conductivity [12]. The operation of a fuel cell at such high temperatures poses major problems related to the stability of the electrolyte-electrolyte interface and selection of the expensive bipolar-plate material [12]. Additionally, for portable (intermittent) power applications, lower-temperature operation is typically favored as it enables rapid start-up and minimizes stress due to thermal cycling [13]. As a result, discovery of potentially new solid electrolytes that can maintain relatively high oxide-ion conductivity at the intermediate temperatures has been an active area of research [14-18]. To this end, several new solid electrolytes have been proposed, such as gadolinium-doped ceria, $\text{La}_{0.9}\text{Sr}_{0.1}\text{Ga}_{0.8}\text{Mg}_{0.2}\text{O}_{2.85}$, $\text{Bi}_2\text{V}_{0.9}\text{Cu}_{0.1}\text{O}_{5.35}$, and Si- and Ge-oxyapatites. Among these potential solid electrolytes Ge-oxyapatites are a relatively new class of materials. Recently Léon-Reina et al. [15] identified high oxide-ion conductivity in Al-doped Ge-oxyapatite with the oxide-ion conductivity comparable to that of gadolinium-doped ceria (see Figure 1). High conductivity coupled with the high ionic transport number in a wide range of oxygen partial pressure (0.21 to 10^{-20} atm) make these Ge-oxyapatite materials attractive for intermediate-temperature, solid-oxide fuel cell application. In contrast with the conventional fluorite and perovskite oxide-ion conductors, where the oxide-ion transport is mediated by vacancy-hopping mechanism, high oxide-ion conductivity in Ge-oxyapatites are interstitial oxygen atoms [16]. The objective of this work is to identify potentially new materials with apatite crystal structure that have high oxide-ion conductivity at the intermediate temperature regime through data mining methods. In particular, we have focused on Ge-oxyapatites because of their superior oxide-ion conducting properties.

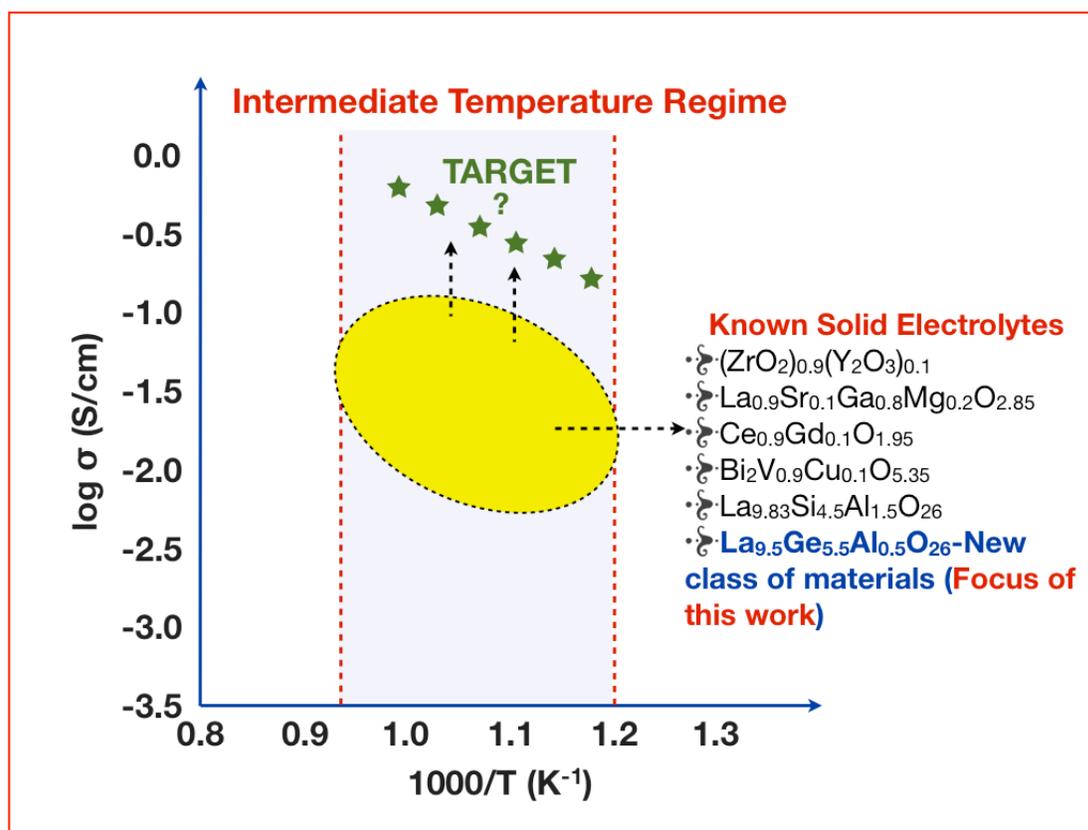


Figure 1. The objective of this research is to identify potentially new solid electrolytes that have high oxide-ion conductivity (represented as TARGET) in the intermediate temperature regime. Only a few solid electrolytes are known that have high oxide-ion conductivity. Ge-oxyapatites (e.g., $\text{La}_{9.5}\text{Ge}_{5.5}\text{Al}_{0.5}\text{O}_{26}$) are a relatively new class of materials with excellent properties for fast ion-conduction. The focus of this work is to develop a computational approach based on data-mining methods for accelerated exploration of the complex Ge-oxyapatite chemical search space in order to rapidly discover potentially new solid electrolytes with oxide-ion properties nearing the TARGET limit. In this paper, we specifically address the challenge associated with the prediction of crystal structure of these complex materials.

Ge-oxyapatites are conveniently described by the general formula $\text{Ln}_{9.33+x}\text{Ge}_6\text{O}_{26+3x/2}$, where the *Ln-sites* usually accommodate larger divalent (Ca^{2+} , Sr^{2+} , Pb^{2+} , Ba^{2+} etc.) and trivalent (Y^{3+} , La^{3+} , Ce^{3+} , Nd^{3+} , Sm^{3+} , Dy^{3+} etc.) cations. Smaller 4^+ , 5^+ and 6^+ metals and metalloids (Ge^{4+} , Si^{4+} , P^{5+} , As^{5+} , V^{5+} , W^{6+} , etc.) fill the *Ge-site*. The complex chemical search space of apatite crystal chemistry is shown in Figure 2. Previous experimental studies have shown that samples with x up to 0.67 can be prepared with single-phase apatite structure [17]. The crystal structures of Ge-apatites play a major role in oxide-ion conduction. Two competing design criteria must be satisfied for high oxide-ion conductivity in Ge-apatites [18]: (1) high concentration of O_2 -interstitials in the lattice and (2) hexagonal crystal structure with $\text{P6}_3/\text{m}$ space group. From

experiments, however, it is known that as the concentration of O_2 ions increases in the unit cell, the crystal structure changes its symmetry from hexagonal to triclinic. Triclinic structure enhances defect trapping; as a result, the oxide-ion conductivity is drastically reduced. The challenge here is to identify potentially new chemistries that help achieve high oxygen contents (in the form of interstitial defects) in the Ge-apatite lattice, while still maintaining hexagonal symmetry. Analogous to the knowledge base of multicomponent materials as mentioned earlier, in the case of apatite structures our existing knowledge (or data) base is relatively small (see Figure 2). Our initial experimental survey has indicated that only about 18% of the total permutation is experimentally explored. Our goal is to explore the remaining 82% of the search space using the data mining. From the initial limited dataset, we propose new materials using information entropy concepts. Shannon entropy metrics coupled with recursive partitioning method is used to discover the specific combination of dominant descriptors that govern the apatite phase stability. From this we extract chemical design rules that allow us to systematically develop relationships resulting in identifying in a quantitative fashion the exact role of specific combination of materials descriptors that govern the stability of a chemical compound.

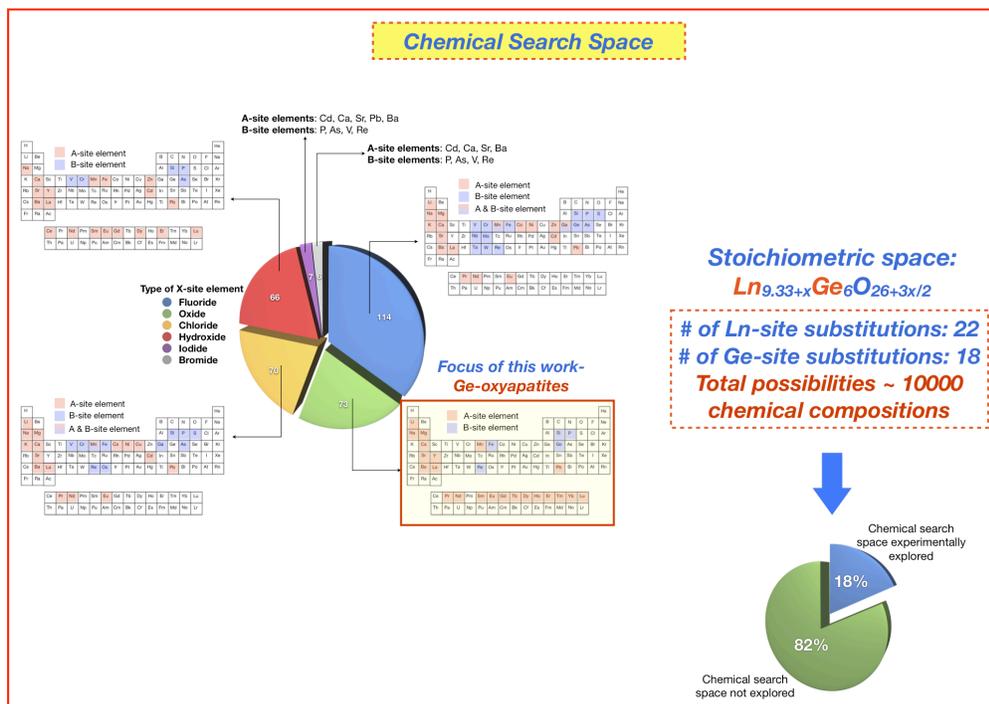


Figure 2. Schematic of the search space of apatite crystal chemistry. Apatites are described by the general formula $A^I_4A^{II}_6(BO_4)_6X_2$, in connection with the contents of a hexagonal unit cell of the space group $P6_3/m$ where A^I and A^{II} are distinct crystallographic sites that usually accommodate larger divalent (Ca^{2+} , Sr^{2+} , Pb^{2+} , Ba^{2+} , etc.), monovalent (Na^+ , Li^+ etc.) and trivalent (Y^{3+} , La^{3+} , Ce^{3+} , Nd^{3+} , Sm^{3+} , Dy^{3+} etc.) cations. B cation sites are filled by smaller 3^+ , 4^+ , 5^+ , 6^+ , and 7^+ metals and metalloids (P^{5+} , As^{5+} , V^{5+} , Si^{4+} , etc.), and the X anion site is occupied by halides (F^- , Cl^- , Br^- , I^-), hydroxyl or oxide ions. In this work, we focus on Ge-oxyapatites, which can be described with the general formula $Ln_{9.33+x}Ge_6O_{26+3x/2}$. In the Ge-oxyapatite stoichiometric space, there are 22 elements that can occupy Ln-site and 18 elements that can occupy Ge-site. Overall, we have more than 10,000 chemical combinations possible. Of which only about 18% have been experimentally explored.

3. Mathematical preliminaries

Recursive-partitioning [19, 20] is an exploratory data analysis technique where we construct models by successively splitting a large dataset into increasingly homogeneous subsets until it is not feasible to continue, based on some well-defined constraints. The output from a recursive partitioning analysis is a dendrogram (or a decision tree) with branches grown on each node to classify the crystal symmetry of Ge-apatites. Let D be the database that contains 77 $Ln_{9.33+x}Ge_6O_{26+3x/2}$ compounds, ten structure governing factors, and a label that explains whether a particular $Ln_{9.33+x}Ge_6O_{26+3x/2}$ compound will have single-phase hexagonal crystal structure or single-phase triclinic crystal structure or not a single-phase apatite region. A total of 49 compounds have single-phase hexagonal crystal structure, 21 compounds have single-phase triclinic structure, and 7 compounds do not have single-phase apatite structure. The expected information required to classify an $Ln_{9.33+x}Ge_6O_{26+3x/2}$ compound solely based on its proportion in the database D is given by the Shannon entropy $H(D)$:

$$H(D) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (1)$$

where p_i is the probability that an arbitrary $Ln_{9.33+x}Ge_6O_{26+3x/2}$ compound in “ D ” belongs to single-phase hexagonal crystal structure or single-phase triclinic crystal structure or not a single-phase apatite region and i is an integer. The unit of entropy is in bits. Now suppose the $Ln_{9.33+x}Ge_6O_{26+3x/2}$ compounds in D are to be partitioned by using some structure-governing factor A . Then the amount of information we would still need (after partitioning) for classification is quantified by

$$H_A(D) = \sum_j \left| \frac{D_j}{D} \right| \times \left(\sum_{i=1}^2 p_j \log_2(p_j) \right) \quad (2)$$

The factor $\left| \frac{D_j}{D} \right|$ acts as the weight of the j th partition. In this work, we have considered only binary split, which implies $j=2$. The smaller the expected information required, the greater the purity of partition. The next step is to compute the “gain ratio,” defined as the ratio of the difference between the original information requirement, $H(D)$ and the new requirement that is obtained after partitioning on A , $H_A(D)$ to the potential information generated by splitting the database D on attribute A .

$$\text{Gain ratio } (A) = \frac{H(D) - H_A(D)}{\sum_j \left| \frac{D_j}{D} \right| \log_2 \left(\left| \frac{D_j}{D} \right| \right)} \quad (3)$$

This process of calculating the gain ratio is repeated for all 13 structure-governing factors and the structure-governing factor with the maximum gain ratio is selected as the splitting attribute. The whole cycle is repeated recursively until it is not feasible to continue splitting the database D based on the stopping criterion (the final node should not contain less than three $Ln_{9.33+x}Ge_6O_{26+3x/2}$ compounds). In order to avoid overfitting, a postpruning operation is

employed. The outcome of recursive partitioning is a dendrogram (or a tree diagram), which captures the physics governing the formation of stable $\text{Ln}_{9.33+x}\text{Ge}_6\text{O}_{26+3x/2}$ apatite compounds.

4. Developing informatics-based crystal chemistry design rules

We begin our approach by considering the materials that we already know from experiments (and hence, by definition, are stable) for which we have a priori information on the concentration of O_2 defects in the lattice and its corresponding crystal structure (our knowledge-base) [14-18]. The key question is, What makes these material chemistries behave the way they do? Our goal is to identify other unexplored material chemistries that may have a similar behavior. Normally electronic structure calculations are employed to address this challenge. Instead, we introduce an alternative approach driven by a data-mining method for accelerated identification of potentially new chemistries satisfying the design constrains. The logic of our approach is described here:

- We characterize the materials by identifying discrete scalar descriptors associated with crystal structure, electronic structure, and crystal chemistry, as shown in Figure 3.

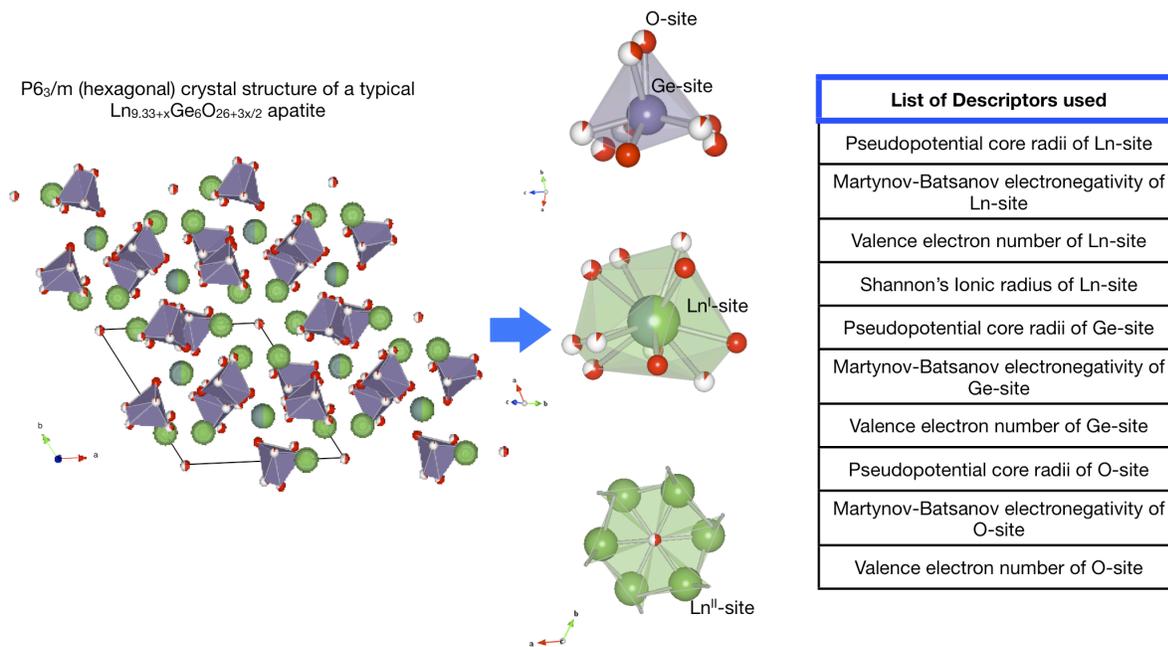


Figure 3. Crystal structure of a typical hexagonal apatite with $\text{P6}_3/\text{m}$ crystal symmetry. The complex crystal structure has two distinct Ln-sites (shown as Ln^{I} and Ln^{II}) and the tetrahedral Ge-site. We describe each Ge-oxyapatite compound in our dataset using the descriptors shown in the right. We employed pseudopotential core radii sum, Martynov-Batsanov electronegativity, valence electron number, and Shannon's ionic radii descriptors. It is well known that these descriptors play an important role in structural determination. For more detailed information on significance of these descriptors, refer to [21].

- Through data mining the high-dimensional knowledge-base, we extract key electronic structure and crystal chemistry features that govern the structural stability.

- The knowledge is extracted in the form of “if ... then” rules that quantitatively identify the exact role of the key materials attributes in describing the crystal structure and properties. The rules are typically visualized in the form of a dendrogram, as shown in Figure 4.
- Using the design rules obtained from the dendrogram, we search the periodic table and virtually add new, unexplored chemistries at various crystallographic sites in the Geapatite lattice. Then we follow the dendrogram, which gives us the crystal structure classification of the new and untested chemical composition.
- This screening strategy down-selects from the initial vast chemical search space a few potential candidates with the required structure, and these new materials are suggested for experimental validation and theoretical modeling.

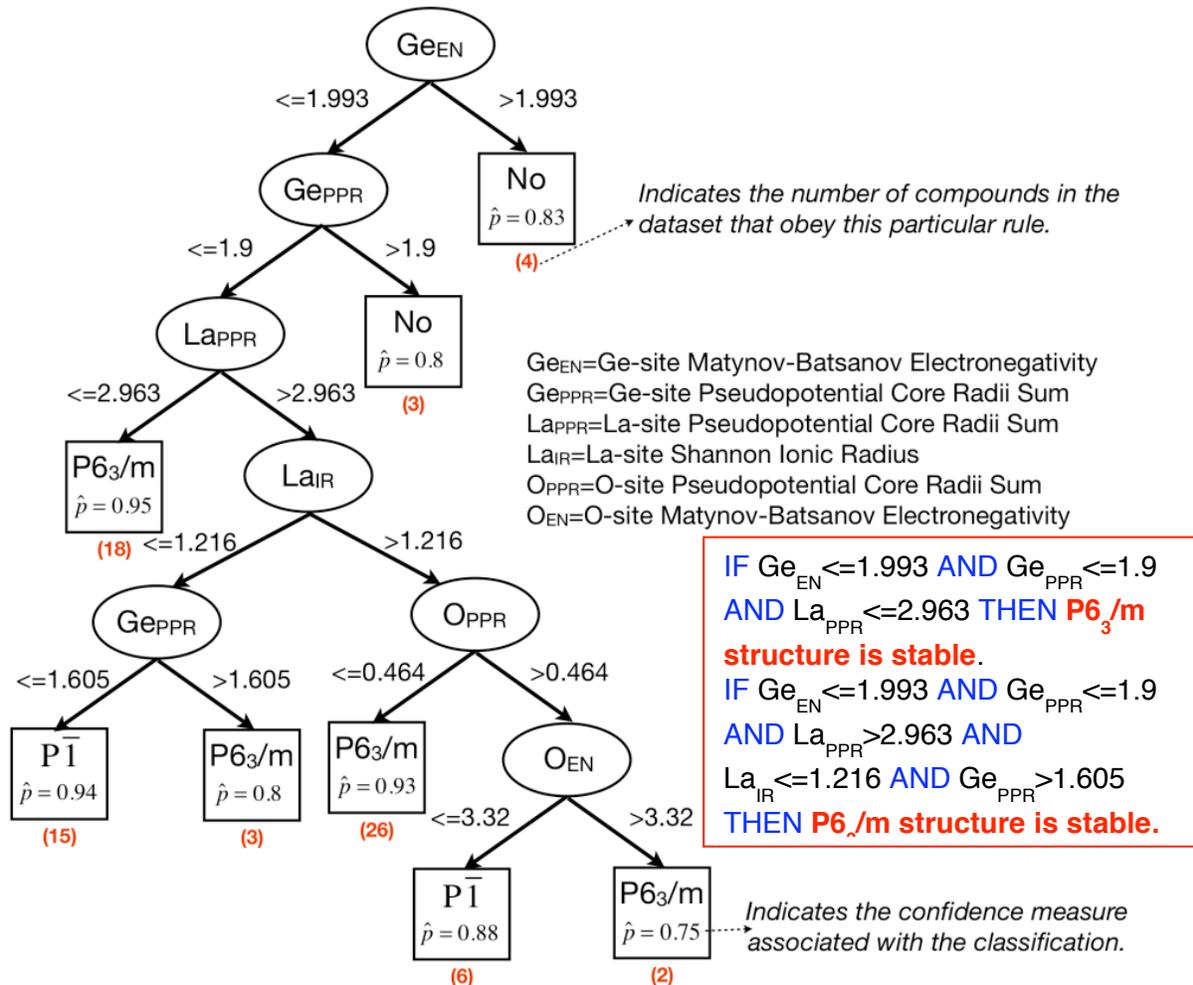


Figure 4. Dendrogram representing the classification model obtained from recursive partitioning. From the dendrogram we extract if ... and ... then design rules (as shown above) and use these rules for classifying the crystal structure of new and untested chemistries.

While the dendrogram provides a guide to the development of the design rules, we can represent this information in the context of how much confidence in the existing knowledge space we have at each stage to make a decision. An alternative visualization schema is provided below that

track the evolution of our information database. This permits us to track which aspect of the electronic structure and its impact on bond geometry is playing a dominant role during the evolution of the design rules. The evolution of information and its physical meaning is discussed in Figures 5–7. The evolution of information discussed in these figures must be interpreted in conjunction with the dendrogram shown in Figure 4.

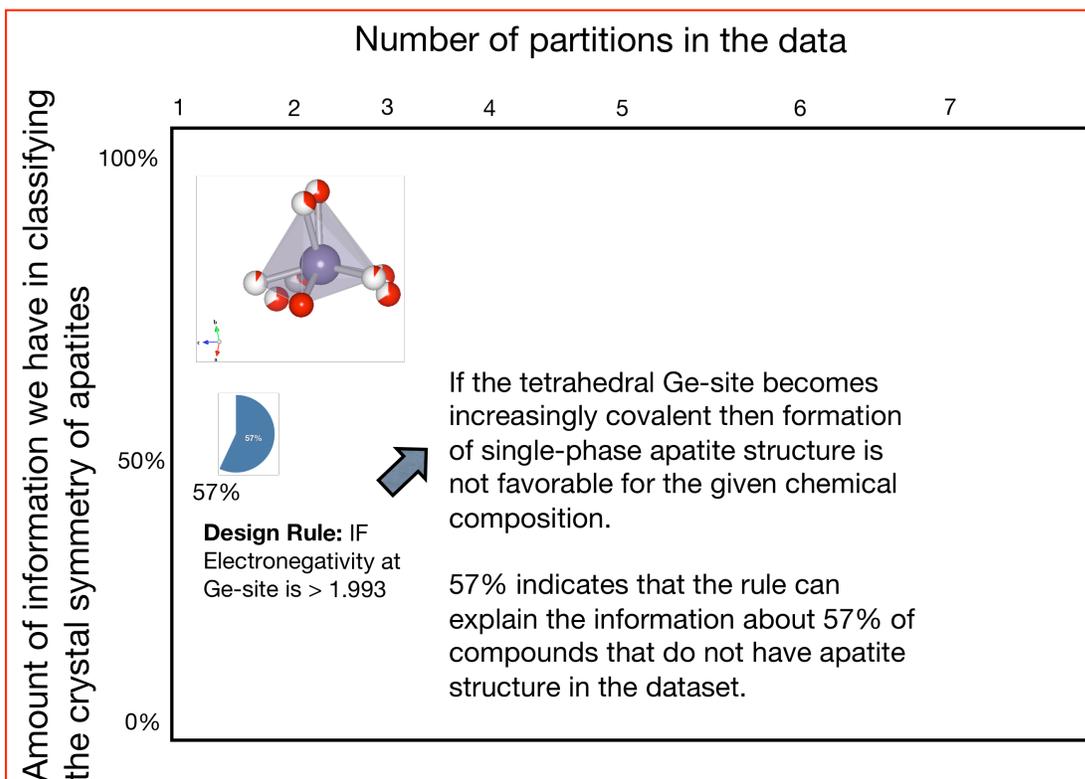


Figure 5. The first design rule states that if the electronegativity at Ge-site is greater than 1.993, then we do not get single-phase apatite structure. Since electronegativity can be correlated with the charge distribution, this rule indicates that if the tetrahedral Ge-site becomes increasingly covalent then obtaining single-phase apatite structure is unfavorable. The value 57% indicates that this rule can explain the information about 57% of compounds that do not have single-phase apatite structure in the dataset. As noted in Section 3, 7 compounds in our dataset do not have single-phase apatite structure. This rule captures the behavior of 4 of 7 compounds in our dataset, or 57%. This information can also be obtained from the decision tree (shown in Figure 4) identified inside the parentheses.

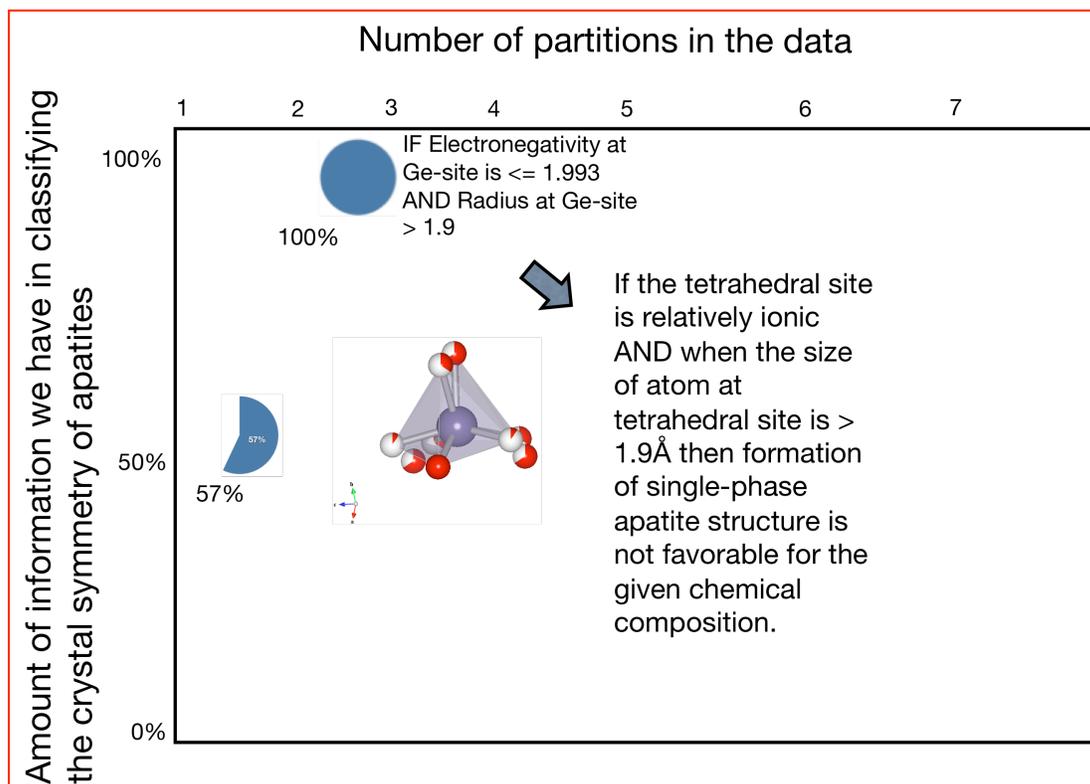


Figure 6. The second design rule suggests that if the electronegativity at Ge-site is less than or equal to 1.993 and the pseudopotential radius at Ge-site is greater than 1.9, then obtaining single-phase apatite structure is not favorable. Rabe et al. [21] have shown that when the electronegativity is combined with pseudopotential radius, then they correlate with the electron charge density information. As a result, this rule quantitatively identifies the key role of electronic structure in impacting the stability of single-phase region in Ge-oxyapatites. The two design rules (Figure 5 and Figure 6) together are sufficient to explain why certain chemistries with the stoichiometry $\text{La}_{10-y}\text{A}_y(\text{Ge}_x\text{B}_{1-x}\text{O}_4)_6\text{O}_2$ do not have single-phase apatite structure. The rule also brings out the significance of the electronic structure of the tetrahedral unit in impacting the phase stability of apatites.

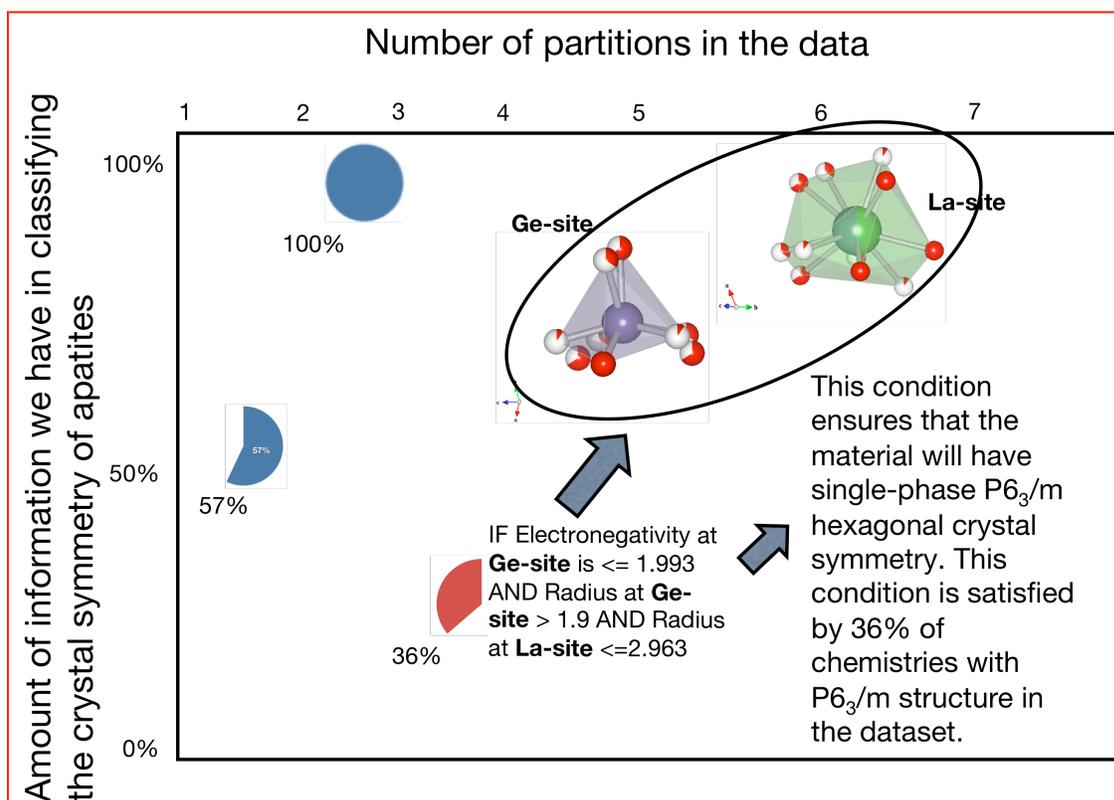


Figure 7. The third design rule states that if the electronegativity at Ge-site is less than or equal to 1.993 and the pseudopotential radius at the Ge-site is greater than 1.9 and the pseudopotential radius at La-site is 2.963, then the chemical composition will have single-phase $P6_3/m$ (hexagonal) crystal symmetry. This condition is satisfied by 36% of chemistries with $P6_3/m$ structure in the dataset. This rule brings out the significance of the tetrahedral and the LnO_6 -structural unit in impacting the phase stability of $P6_3/m$ symmetry.

Similarly, we can further explore the dendrogram and investigate the key physical reason behind the structural stability of Ge-oxyapatites. The visualization scheme discussed in Figures 5–7 refines our understanding of the complex Ge-oxyapatite crystal chemistry, which has to be further augmented by performing quantum mechanical calculations and experimental studies on chemical compositions that are classified to have $P6_3/m$ crystal structure by the dendrogram. The dendrogram thus allows for accelerated screening of novel chemical compositions to guide highly targeted experimental and detailed theoretical studies of limited compositional regions. Thus, from a set of very limited initial dataset (18%) we identify the best solutions by partitioning the descriptor space. The Shannon entropy metric captures the association between the sequence of descriptors and the structural stability of Ge-oxyapatites. We track the evolution of information via visualizing the dendrogram and suggest a few new compounds with the target $P6_3/m$ crystal structure by virtually exploring the 82% chemical search space. This is schematically shown in Figure 8.

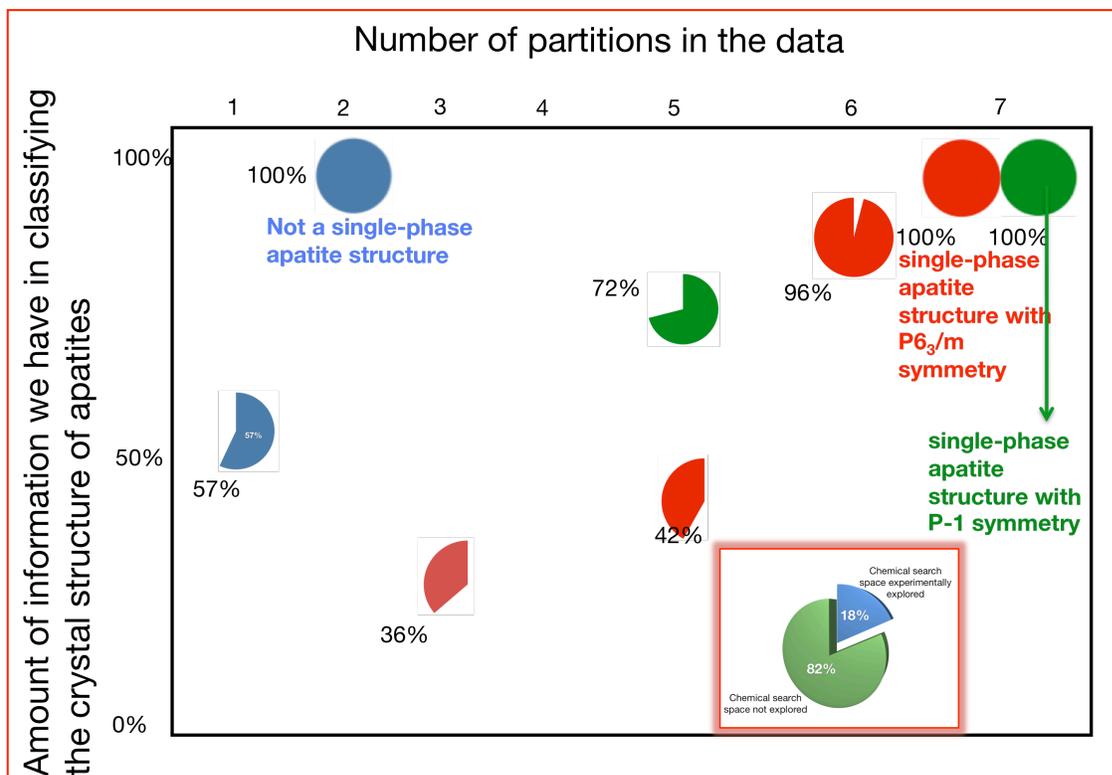


Figure 8. By combining the design rules deduced from dendrogram with the visualization scheme, we clearly understand the evolution of the information in the context of how much confidence of the existing knowledge space we have at each stage to make a decision. Thus by employing 18% of known experimental information, we can rapidly explore the unknown 82% chemical search space. Then we screen for novel chemical compositions to identify only those chemical compositions that form $P6_3/m$ crystal structure. This screening strategy down-selects from the initial vast chemical search space a few potential candidates with the required structure, and these new materials are suggested for further experimental validation and detailed theoretical modeling.

5. Summary

In materials science, we encounter large quantities of sparse datasets. In this paper, we have shown how we can harness the information in smaller datasets for mapping and tracking structure-chemistry correlations in complex materials for expediting new materials discovery. Normally, crystallography is associated with visualization of crystal structure in the physically conceivable three dimensions. However, the success of our approach lies in moving beyond the three dimensions, where we employ a high-dimensional dataset using the discrete scalar descriptors associated with geometric and electronic structure attributes. The information entropy coupled with a recursive partitioning strategy helps in partitioning the high-dimensional descriptor space. The partitioning results in the extraction of chemical design rules, which suggests new materials. The representation of information in the context of how much confidence of the existing knowledge space we have at each stage to make a decision helps in linking uncertainty with materials discovery. We demonstrated the utility of our approach by

building chemical design rules for classify the crystal structure of complex Ge-oxyapatites for intermediate temperature solid oxide fuel cell applications.

Acknowledgments

We acknowledge support from the National Science Foundation:, NSF-ARI Program: CMMI 09-389018; NSF-CDI Type II program: grant no. PHY 09-41576 and NSF-AF grant no. CCF09-17202, NSF-CRI grant no. 07-51157; the Air Force Office of Scientific Research, grant no FA9550-08-1-0316;; and Army Research Office grant no. W911NF-10-0397. KR also acknowledges support from Iowa State University through the Wilkinson Professorship in Interdisciplinary Engineering

References

- [1] Maddox J. 1988. Crystals from first principles, *Nature*, **335**, p. 201.
- [2] Curtarolo S, Morgan D, Persson K, Rodgers J, and Ceder, G. 2003. Predicting crystal structures with data mining of quantum calculations, *Phys. Rev. Lett.*, **91**, pp. 135–503.
- [3] Hart G L W, Blum V, Walorski M J, Zunger A. 2005. Evolutionary approach for determining first principles Hamiltonians, *Nat. Mater.*, **4**, pp. 391-394.
- [4] Pettifor D. 1995. *Bonding and Structure of Molecules and Solids*, Oxford, Oxford University Press.
- [5] Kanatzidis, M G, et al. 2008. Report from the third workshop on future directions of solid-state chemistry: The status of solid-state chemistry and its impact in the physical sciences, *Prog. Solid State Chem.*, **36**, pp. 1-133.
- [6] Mackay A L. 1986. Generalized crystallography science on form. In *Proc. of the 1st Int. Symposium on Form, Tsukuba, Japan*, eds. Y. Kato, R. Takaki, and J. Toriwaki, Tokyo, Japan: KTK Scientific Publishers, pp. 615–620.
- [7] Mackay A L. 1995. Generalized crystallography, *J. Mol. Struct.*, **336**, pp. 293-303.
- [8] Mackay A L. 2002. Generalized crystallography, *Struct. Chem.*, **13**, pp. 215-220.
- [9] Balachandran P V, and Rajan K. 2011. Identifying the “inorganic gene” for high-temperature piezoelectric perovskites through statistical learning, *Proc. R. Soc. A* (in press) doi: 10.1098/rspa.2010.0543.
- [10] Tolchard J R, Slater P R, and Islam, M S. 2007. Insight into doping effects in apatite silicate ionic conductors, *Adv. Funct. Mater.*, **17**, pp. 2564-2571.
- [11] Huijsmans J P P, van Berkel F P F, and Christie G M. 1998. Intermediate temperature SOFC – a promise for the 21st century, *J. Power Sources*, **71**, pp. 107-110.
- [12] Steele B C H, and Heinzl A. 2001. Materials for fuel-cell technologies, *Nature*, **414**, pp. 345-352.
- [13] Haile S M. 2003. Fuel cell materials and components, *Acta Mat.*, **51**, pp. 5981-6000.
- [14] Orera A. et al. 2011. Apatite germinates doped with tungsten: synthesis, structure and conductivity, *Dalton Trans.*, **40**, pp. 3903-3908.
- [15] Léon-Reina L., et al. 2005. High oxide ion conductivity in Al-doped germanate oxyapatite, *Chem. Mater.*, **17**, pp. 596-600.
- [16] Malavasi L., Fischer C A J, and Islam M S. 2010. Oxide-ion and proton conducting electrolyte materials for clean energy applications: structural and mechanistic features, *Chem. Soc. Rev.*, **39**, pp. 4370-4387.

- [17] Léon-Reina L, et al. 2003. Crystal chemistry and oxide ion conductivity in the lanthanum oxygemanate apatite series, *Chem. Mater.*, **4**, pp. 2099–2108.
- [18] Orera A, et al. 2011. Strategies for the optimization of the oxide ion conductivities of apatite-type germinates, *Fuel Cells*, **11**, pp. 10-16.
- [19] Han J. and Kamber M. 2006 *Data mining: concepts and techniques*, 2nd ed., pp. 291–310. Amsterdam, The Netherlands: Elsevier.
- [20] Hall M, Frank E, Holmes G, Pfahringer B., Reutemann P, and Witten I. H. 2009. The WEKA Data Mining software: an update, *SIGKDD Exploration* , **11** pp. 10–18.
- [21] Rabe K M, Phillips J C, Villars P, Brown I D. 1992. Global multinary structural chemistry of stable quasicrystals, high Tc ferroelectrics and high Tc superconductors, *Phys. Rev. B*, **45**, pp. 7650-7676.